Antonio Fernández-Caballero
María Gracia Manzano Arjona
Enrique Alonso González
Sergio Miguel Tomé (Eds.)

Una Perspectiva de la Inteligencia Artificial en su 50 Aniversario

Campus Multidisciplinar en Percepción e Inteligencia, CMPI-2006
Albacete, España, 10-14 de Julio del 2006
Actas, Volumen I

Universidad de Castilla-La Mancha

Departamento de Sistemas Informáticos

Impreso en España. Printed in Spain.

ISBN 84-689-9560-6 (Obra completa)
ISBN 84-689-9561-4 (Volumen I)

Depósito Legal: AB-314-2006

Imprime: Gráficas Quintanilla. La Roda

Diseño de la cubierta: UGSC (Unidad de Gestión Sociocultural)

Presentación

El 31 de Agosto 1955, *J. McCarthy* (Dartmouth College, New Hampshire), *M.L. Minsky* (Harvard University), *N. Rochester* (I.B.M. Corporation) y C.E. Shannon (Bell Telephone Laboratories) lanzaron una propuesta para reunir en el verano de 1956 a un grupo de investigadores que quisieran trabajar sobre la conjetura de que cada aspecto del aprendizaje y cada característica de la inteligencia podían ser tan precisamente descritos que se podían crear máquinas que las simularan. El encuentro, celebrado en 1956 y ahora conocido como la conferencia de Dartmouth, se llevó a cabo con tal éxito que el evento acuñó el término *Inteligencia Artificial* y con él una nueva área científica de conocimiento. En el año 2006 se cumplen cincuenta años de la Conferencia de Dartmouth. Pero a pesar del tiempo transcurrido, el problema de encontrar las minuciosas descripciones de las características del cerebro y de la mente que fue mencionado en la propuesta de 1955 sigue tan vigente hoy, como ayer, a pesar del variado abanico de ciencias que lo abordan y estudian.

Albacete (España) ha sido en la semana del 10 al 14 de Julio la sede del evento internacional más importante en lengua castellana con el *Campus Multidisciplinar en Percepción e Inteligencia, CMPI-2006*. El Campus Multidisciplinar en Percepción e Inteligencia 2006 es un evento internacional en el que investigadores de diversas áreas relacionadas con la Percepción y la Inteligencia se encontrarán del 10 al 14 de Julio en el Campus Universitario de Albacete con el ánimo de recuperar el espíritu entusiasta de aquellos primeros días de la Inteligencia Artificial. En nuestra intención está el objetivo de crear un ambiente heterogéneo formado por especialistas de diversas áreas, cómo la Inteligencia Artificial, la Neurobiología, la Psicología, la Filosofía, la Lingüística, la Lógica, la Computación,, con el fin de intercambiar los conocimientos básicos de las diferentes áreas y de poner en contacto investigadores de los diferentes campos. El facilitar la creación de colaboraciones e investigaciones multidisciplinares es un objetivo prioritario de la propuesta.

El *Congreso Multidisciplinar en Percepción e Inteligencia*, que ha dado lugar a esta publicación, se engloba como parte fundamental en el Campus Multidisciplinar sobre Percepción e Inteligencia. Este Congreso Multidisciplinar en Percepción e Inteligencia va dirigida a todas aquellas personas que tengan interés por conocer qué es la Percepción y qué es la Inteligencia, vistas ambas desde una perspectiva claramente multidisciplinar. El Congreso Multidisciplinar contará con la presencia de destacados especialistas del campo de la investigación. Todos ellos, así, y desde su propia experiencia, podrán proporcionar a los asistentes una visión muy clara del estado actual de las distintas ciencias que se ocupan de la Percepción y la Inteligencia. Estas charlas invitadas o tutoriales complementan a la perfección las ponencias que se impartirán por las mañanas durante la Escuela de Verano sobre Percepción e Inteligencia.

El Campus Multidisciplinar en Percepción e Inteligencia ha contado también con la *Escuela de Verano en Percepción e Inteligencia: 50 Aniversario de la Inteligencia Artificial*, que se ha ofertado en el seno de la XIX Edición de Cursos de Verano de la Universidad de Castilla-La Mancha, y que se ha correspondido con la XVI Escuela de Verano del Departamento de Sistemas Informáticos. Pensada fundamentalmente para los alumnos de la Universidad de Castilla-La Mancha del Campus de Albacete, la Escuela de Verano sobre Percepción e Inteligencia ha cubierto aspectos de gran interés para las carreras de Informática, Medicina, Humanidades y Magisterio. Las clases magistrales de la Escuela de Verano han estado a cargo de importantes y reconocidos investigadores a nivel internacional. Todos ellos, así, y desde su propia experiencia, han proporcionado a los asistentes una visión muy clara del estado actual de las distintas ciencias que se ocupan de la Percepción y la Inteligencia.

Fruto de las contribuciones más importantes del evento han nacido dos libros[1]. El primero de ellos, formado por dos volúmenes, denominado *Una Perspectiva de la Inteligencia Artificial en su 50 Aniversario. Actas del Campus Multidisciplinar en Percepción e Inteligencia, CMPI-2006*, es el que tiene en sus manos. Contiene las contribuciones de los congresistas expuestas oralmente o presentadas como pósters en el Congreso. El otro, llamado *50 Años de la Inteligencia Artificial. XVI Escuela de Verano de Informática*, recoge las ponencias invitadas de prestigiosos investigadores que han asistido al Campus Multidisciplinar en Percepción e Inteligencia.

Julio del 2006

Antonio Fernández-Caballero
María Gracia Manzano Arjona
Enrique Alonso González
Sergio Miguel Tomé
Comité Organizador CMPI-2006

[1] La publicación de estos libros ha sido financiada en parte por la Acción Complementaria del Ministerio de Educación y Ciencia TIN2005-25897-E y la Acción Especial de la Junta de Comunidades de Castilla-La Mancha AEB06-023.

Organización

El congreso internacional *50 Años de la Inteligencia Artificial: Campus Multidisciplinar en Percepción e Inteligencia, CMPI-2006* ha sido organizado por el Departamento de Sistemas Informáticos (DSI) y el Instituto de Investigación en Informática de Albacete (I3A) de la Universidad de Castilla-La Mancha (UCLM) en cooperación con el Parque Científico y Tecnológico de Albacete (PCyTA) y el Excmo. Ayuntamiento de Albacete.

Comité Organizador

Antonio Fernández-Caballero	(Universidad de Castilla-La Mancha)
María Gracia Manzano Arjona	(Universidad de Salamanca)
Enrique Alonso González	(Universidad Autónoma de Madrid)
Sergio Miguel Tomé	(Universidad de Castilla-La Mancha)

Comité Local

De la Ossa Jiménez, Luis	(Universidad de Castilla-La Mancha)
Díaz Delgado, Carmen	(Universidad de Castilla-La Mancha)
Fernández Graciani, Miguel Angel	(Universidad de Castilla-La Mancha)
Flores Gallego, María Julia	(Universidad de Castilla-La Mancha)
García Varea, Ismael	(Universidad de Castilla-La Mancha)
Gómez Quesada, Francisco J.	(Universidad de Castilla-La Mancha)
López Bonal, María Teresa	(Universidad de Castilla-La Mancha)
López Valles, José María	(Universidad de Castilla-La Mancha)
Mateo Cerdán, Juan Luis	(Universidad de Castilla-La Mancha)
Miranda Alonso, Tomás	(Universidad de Castilla-La Mancha)
Parreño Torres, Francisco	(Universidad de Castilla-La Mancha)
Ponce Sáez, Antonio	(Universidad de Castilla-La Mancha)

Comité de Programa

Adán Oliver, Antonio	Dept. de Ingeniería Eléctrica, Electrónica y Automática - Universidad de Castilla-La Mancha en Ciudad Real (ES)
Alvarez Sánchez, José Ramón	Dept. de Inteligencia Artificial - Universidad Nacional de Educación a Distancia (ES)
Arce Michel, Javier	Sociedad Iberoamericana de Psicología, Bolivia - SIAPSI (BO)
Areces, Carlos	Langue et Dialogue - Laboratoire Lorrain de Recherche en Informatique et ses Applications (FR)
Armengol i Voltas, Eva	Institut d'Investigació en Intel.ligència Artificial - Centro Superior de Investigaciones Científicas (ES)

Armero San José, Julio	Dept. de Lógica, Historia y Filosofía de la Ciencia - Universidad Nacional de Educación a Distancia (ES)
Augusto, Juan Carlos	School of Computing and Mathematics - University of Ulster at Jordanstown (UK)
Barber Sanchís, Federico	Dept. de Sistemas Informáticos y Computación - Universitat Politècnica de Valencia (ES)
Barro Ameneiro, Senén	Dept. de Electrónica y Computación - Universidade de Santiago de Compostela (ES)
Bayro-Corrochano, Eduardo José	Computer Science - CINVESTAV, Guadalajara (MX)
Bel Enguix, Gemma	Research Group in Mathematical Linguistics - Universitat Rovira i Virgili (ES)
Blanco Mayor, Aquilino Carmelo	Dept. de Filosofía - Universidad de Castilla-La Mancha en Albacete (ES)
Botella Beviá, Federico	Dept. de Estadística, Matemática e Informática - Universidad Miguel Hernández de Elche (ES)
Bustos Guadaño, Eduardo de	Dept. de Lógica, Historia y Filosofía de la Ciencia - Universidad Nacional de Educación a Distancia (ES)
Caminos Benito, María Elena	Centro Regional de Investigaciones Biomédicas - Universidad de Castilla-La Mancha en Albacete (ES)
Casacuberta Nolla, Francisco	Dept. de Sistemes Informàtics i Computació - Universitat Politècnica de València (ES)
Casals Gelpí, Alicia	Dept. d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial - Universitat Politècnica de Catalunya (ES)
Castejón Costa, Juan Luis	Dept. Sociologia II, Psicología, Comunicación y Didáctica - Universitat d'Alacant (ES)
Chinellato, Eris	Robotic Intelligence Laboratory - Universitat Jaume I (ES)
Cordón García, Oscar	Applications of Fuzzy Logic and Evolutionary Algorithms Research Unit - European Centre for Soft Computing (ES)
Cortés, Ulises	Knowledge Engineering and Machine Learning Group - Universitat Politècnica de Catalunya (ES)
Cuadros-Vargas, Ernesto	Computer Science - Universidad Catolica San Pablo, Arequipa (PE)
Deco, Gustavo	Institució Catalana de Recerça i Estudis Avançats - Universitat Pompeu Fabra (ES)
Del Pobil, Angel Pasqual	Robotic Intelligence Laboratory - Universitat Jaume I (ES)
Delgado García, Ana	Dept. de Inteligencia Artificial - Universidad Nacional de Educación a Distancia (ES)
Díaz Delgado, Carmen	Centro Regional de Investigaciones Biomédicas - Universidad de Castilla-La Mancha en Albacete (ES)
Duro, Richard J.	Dept. de Computación - Universidade da Coruña (ES)
Faríñas del Cerro, Luis	Institut de Recherche en Informatique de Toulouse - Université Paul Sabatier (FR)

Feliú Batlle, Vicente	Dept. de Ingeniería Eléctrica, Electrónica y Automática - Universidad de Castilla-La Mancha en Ciudad Real (ES)
Fernández Graciani, Miguel Angel	Dept. de Sistemas Informáticos - Universidad de Castilla-La Mancha en Albacete (ES)
Fernández Moreno, Luis	Dept. de Lógica y Filosofía de la Ciencia - Universidad Complutense de Madrid (ES)
Fuentes Melero, Luis	Dept. de Psicología Básica y Metodología - Universidad de Murcia (ES)
Gámez Martín, José Antonio	Dept. de Sistemas Informáticos - Universidad de Castilla-La Mancha en Albacete (ES)
García Pupo, Mauro	Sociedad Cubana de Matemática y Computación - Universidad de Holguín (CU)
García Varea, Ismael	Dept. de Sistemas Informáticos - Universidad de Castilla-La Mancha en Albacete (ES)
Garijo, Francisco	Telefónica Investigación y Desarrollo - Telefónica (ES)
Geffner, Hector	Institució Catalana de Recerca i Estudis Avançats - Universitat Pompeu Fabra (ES)
Godó Lacasa, Lluis	Institut d'Investigació en Intel.ligència Artificial - Centro Superior de Investigaciones Científicas (ES)
González López, Pascual	Dept. de Sistemas Informáticos - Universidad de Castilla-La Mancha en Albacete (ES)
Graña Romay, Manuel	Dept. de Ciencias de la Computación e Inteligencia Artificial - Euskal Herriko Unibertsitatea / - Universidad del País Vasco (ES)
Hernández Orallo, José	Dept. de Sistemes Informàtics i Computació - Universitat Politècnica de València (ES)
Herrera Viedma, Enrique	Dept. de Ciencias de la Computación e Inteligencia Artificial - Universidad de Granada (ES)
Huertas Sánchez, M. Antonia	Informática y Multimedia - Universitat Oberta de Catalunya (ES)
Insausti Serrano, Ricardo	Centro Regional de Investigaciones Biomédicas - Universidad de Castilla-La Mancha en Albacete (ES)
Jansana, Ramón	Dept. de Lògica, Història i Filosofia de la Ciència - Universitat de Barcelona (ES)
Jiménez López, María Dolores	Research Group in Mathematical Linguistics - Universitat Rovira i Virgili (ES)
Juiz Gómez, José Manuel	Dept. de Ciencias Médicas - Universidad de Castilla-La Mancha en Albacete (ES)
Kemper Valverde, Nicolás	Laboratorio de Sistemas Inteligentes - Universidad Nacional Autónoma de México (MX)
Latorre Postigo, José Miguel	Dept. de Psicología - Universidad de Castilla-La Mancha en Albacete (ES)
Llinás Riascos, Rodolfo	Department of Physiology and Neuroscience - New York - University Medical Center (USA)
Llopis Borrás, Juan	Centro Regional de Investigaciones Biomédicas - Universidad de Castilla-La Mancha en Albacete (ES)
López Bonal, María Teresa	Dept. de Sistemas Informáticos - Universidad de Castilla-La Mancha en Albacete (ES)

Puelles López, Luis	Dept. de Anatomía Humana y Psicobiología - Universidad de Murcia (ES)
Puerta Callejón, José Miguel	Dept. de Sistemas Informáticos - Universidad de Castilla-La Mancha en Albacete (ES)
Rechea Alberola, Cristina	Dept. de Psicología - Universidad de Castilla-La Mancha en Albacete (ES)
Rodríguez Ladreda, Rosa María	Asociación Andaluza de Filosofía (ES)
Ruiz Shulcloper, José	Centro de Aplicaciones de Tecnologías de Avanzada - CENATAV, Ciudad Cuba (CU)
Sánchez-Andrés, Juan Vicente	Dept. de Fisiología - Universidad de La Laguna (ES)
Sánchez Calle, Ángel	Dept. de Informática, Estadística y Telemática - Universidad Rey Juan Carlos (ES)
Sánchez Cánovas, José	Dept. de Personalitat, Avaluació i Tractaments Psicològics - Universitat de València (ES)
Sánchez Vila, Eduardo	Dept. de Electrónica y Computación - Universidade de Santiago de Compostela (ES)
Sanfeliú Cortés, Alberto	Institut de Robòtica i Informàtica Industrial - Universitat Politècnica de Catalunya (ES)
Solar Fuentes, Mauricio	Dept. de Ingeniería Informática - Universidad de Santiago de Chile (CL)
Silva Mata, Francisco José	Centro de Aplicaciones de Tecnologías de Avanzada - CENATAV, Ciudad Cuba (CU)
Sossa Azuela, Juan Humberto	Centro de Investigación en Computación - Instituto Politécnico Nacional (MX)
Sucar Succar, Luis Enrique	Dept. de Computación - Instituto Tecnológico y de Estudios Superiores de Monterrey (MX)
Taboada Iglesias, María Jesús	Dept. de Electrónica e Computación - Universidade de Santiago de Compostela (ES)
Toro-Alfonso, José	Dept. de Psicología - Universidad de Puerto Rico
Torra, Vicenç	Institut d'Investigació en Intel-ligència Artificial - Centro Superior de Investigaciones Científicas (ES)
Trillas Ruiz, Enric	Dept. de Inteligencia Artificial - Universidad Politécnica de Madrid (ES)
Tudela Garmendia, Pío	Dept. de Psicología Experimental y Fisiología del Comportamiento - Universidad de Granada (ES)
Vega Reñón, Luis	Dept. de Lógica, Historia y Filosofía de la Ciencia - Universidad Nacional de Educación a Distancia (ES)
Verdegay Galdeano, José Luis	Dept. de Ciencias de la Computación e Inteligencia Artificial - Universidad de Granada (ES)
Vidal Ruiz, Enrique	Dept. de Sistemes Informàtics i Computació - Universitat Politècnica de València (ES)

Entidades Organizadoras

Universidad de Castilla-La Mancha
Parque Científico y Tecnológico de Albacete
Excmo. Ayuntamiento de Albacete

Entidades Patrocinadoras

Ministerio de Educación y Ciencia
Junta de Comunidades de Castilla-La Mancha
(Consejería de Educación y Ciencia)
Caja Castilla-La Mancha
Telefónica
Fundación Campollano
Instituto de Investigación en Informática de Albacete
Departamento de Sistemas Informáticos, UCLM
Revista "Mente y Cerebro"
Centro Regional de Investigaciones Biomédicas, UCLM
Excma. Diputación de Albacete

Entidades Colaboradoras

Asociación Española para la Inteligencia Artificial
Asociación Andaluza de Filosofía
Associació Catalana d'Intel·ligència Artificial
Asociación Cubana de Reconocimiento de Patrones
Fundación Española para la Ciencia y la Tecnología
Instituto de Filosofía, CSIC
Instituto de Neurociencias de Castilla y León
Mexican Association for Computer Vision, Neurocomputing and Robotics
Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España
Sociedad Chilena de Ciencia de la Computación
Sociedad Colombiana de Psicología
Sociedad Cubana de Matemática y Computación
Sociedad Española de Filosofía Analítica
Sociedad Española de Neurociencia
Sociedad Española de Psicología Experimental
Sociedad Interamericana de Psicología
Sociedad Mexicana de Ciencia de la Computación
Sociedad Peruana de Computación
Sociedad Venezolana de Filosofía
TECNOCIENCIA

Índice general

VOLUMEN I

SISTEMAS EXPERTOS Y DE AYUDA A LA DECISIÓN

FILOSOFÍA Y MODELOS DE LA MENTE

REDES NEURONALES

COMPUTACIÓN EVOLUTIVA Y ALGORITMOS GENÉTICOS

ROBÓTICA Y SISTEMAS AUTÓNOMOS

SISTEMAS MULTIAGENTE Y ARQUITECTURAS PARA LA INTELIGENCIA ARTIFICIAL

VOLUMEN II

PERCEPCIÓN E INTELIGENCIA BIO-INSPIRADAS

CLASIFICACIÓN Y RECONOCIMIENTO DE PATRONES

RAZONAMIENTO FORMAL

RAZONAMIENTO APROXIMADO Y RAZONAMIENTO BAYESIANO

HEURÍSTICAS Y METAHEURÍSTICAS

TRADUCCIÓN AUTOMÁTICA

TECNOLOGÍAS DE INTERACCIÓN INTELIGENTES

PLANIFICACIÓN Y OPTIMIZACIÓN

Inteligencia artificial frente a inteligencia natural cuando expresamos actitudes

Antonio Jesús Herencia-Leva[1] y Mª Teresa Lamata[2]

[1] Centro de Psicología AARON BECK. 18002. Granada, España
antonio@cpaaronbeck.com
[2] Depto de Ciencias de la Computación e Inteligencia Artificial.
E.T.S de Ingeniería Informática. Universidad de Granada.18071.Granada, España
mtl@decsai.ugr.es

Resumen. El estudio de las preferencias, valores y actitudes de las personas y la forma en que las máquinas podrían reflejar estos sentimientos, es de gran importancia para construir agentes inteligentes. Las personas expresamos nuestras opiniones y actitudes de una forma que todavía no ha podido ser modelada. Nos caracterizamos, entre otras cosas, por se capaces de actuar en situaciones de gran incertidumbre. En este trabajo se presentan los principales tipos de incertidumbre lingüística que se han estudiado y las diferentes alternativas que se han propuesto para su medición y manejo. La teoría de conjuntos difusos a través de la definición de funciones de pertenencia, permite reflejar adecuadamente estos tipos de incertidumbre lingüística. Sin embargo, cuando se trata de estudiar propiedades como las opiniones y las actitudes, para las que no se dispone de un referente físico o psicológico claro, aparecen múltiples problemáticas que todavía no se han sabido responder.

Palabras Clave: incertidumbre, vaguedad, difuso, funciones de pertenencia, medición de actitudes, toma de decisiones, escalabilidad, términos lingüísticos

1 Introducción

A lo largo de la historia de la Inteligencia Artificial (IA) se han defendido diferentes posturas acerca de la relación entre la IA y la inteligencia natural. Esta discusión se ha visto en gran parte limitada y condicionada a la postura reinante en el momento en torno a la relación entre mente y cuerpo. Esta dependencia fue superada inicialmente optando por un posicionamiento conductista: para Turing la inteligencia viene determinada por aquello que se puede hacer y aquello que no se puede hacer; un no-humano se puede considerar inteligente si interactuando con otro humano, un tercero que está observando la conversación, no es capaz de distinguirlos en su habilidad verbal. Esta definición, a pesar de suponer una superación de la postura mentalista, prestaba la limitación de dejar en su propia definición circular el criterio para determinar qué se podía considerar inteligente. Se avanzó en esta postura a partir de la introducción de Dennet del concepto de intencionalidad: una conducta es intencionada cuando forma parte de un patrón de conductas racionales, viniendo la racionalidad dada por la elección de la mejor alternativa a la luz de conocimiento que se posee y

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 1-11, 2006.

de la situación en la que se desarrolla [1].

Son principalmente los intereses, las actitudes, opiniones y valores que tiene una persona, los que dan cierta unidad, coherencia y predecibilidad a su forma de comportarse. No sólo es importante para la IA que una máquina sea capaz de expresar actitudes, sino que sea capaz de reconocerlas en las personas y adaptar su comportamiento a dicha información.

En Ciencias Sociales, este ámbito de estudio se conoce como habilidades sociales [2]. Las habilidades sociales no son meramente un conjunto de protocolos de forma de comportarse que una máquina podría aplicar teniendo en cuenta únicamente claves estimulares (como tipo de persona y situación) y un conjunto de criterios a maximizar a corto y largo plazo. Uno de los aspectos más importantes es la asertividad, vinculada a que a partir del contenido verbal y no verbal del comportamiento del agente, se puedan inferir sus valores, actitudes, intereses y objetivos en la vida. Esto es lo que realmente da autenticidad, coherencia y racionalidad a su forma de comportarse.

A nivel verbal, las personas nos caracterizamos por utilizar un número relativamente reducido de términos lingüísticos, adecuando el nivel de información transmitido a las necesidades de la situación, y consiguiendo al mismo tiempo que nuestro mensaje sea correctamente entendido [3]. Se habla de que las personas somos capaces de manejar adecuadamente la incertidumbre que rodea al uso e interpretación de los términos lingüísticos en nuestras conversaciones diarias.

La Teoría de Conjuntos Difusos (TCD) de Zadeh [4] se ha mostrado especialmente útil a la hora de representar términos lingüísticos y operar con éstos términos bajo condiciones de incertidumbre.

Sin embargo, la construcción de conjuntos difusos para términos lingüísticos como agresivo, bonito, atractivo y otros para los que no se dispone de un referente, presentan dificultades añadidas. En la mayor parte de los trabajos sobre conjuntos difusos se hace referencia a términos lingüísticos como alto, para los que se dispone como referente la longitud medida en centímetros o metros. Cuando se aborda el estudio de las preferencias hacia objetos sociales el principal problema que aparece es la ausencia de un referente claro que permita construir los conjuntos difusos y de esta forma poder modelar cómo las personas manejamos la incertidumbre lingüística.

En este trabajo se presenta una propuesta para resolver este problema: inicialmente se construye un referente a partir del grado de preferencia que tienen los sujetos hacia un conjunto de objetos y posteriormente se construyen los conjuntos difusos. Este proceso se puede hacer iterativo hasta que se consiga representar adecuadamente la información que se dispone.

Este trabajo se organiza como sigue: en la sección 2 se hace una revisión del estudio de diferentes tipos de incertidumbre lingüística, centrándose esta revisión en la distinción entre los conceptos de vaguedad, ambigüedad, generalidad y difuso. En la sección 3 se revisa la metodología para construir conjuntos difusos y medir funciones de pertenencia. Por último, se finaliza exponiendo varias conclusiones y líneas de investigación futuras.

2 Cuantificación de la vaguedad

La vaguedad de un término lingüístico se relaciona con el grado en que tal término es aplicable a una serie de objetos [5]. Muy vinculado a la vaguedad se encuentra la generalidad, ambos reflejan diferentes aspectos del hecho de que un mismo término sea aplicado a diferentes objetos. La generalidad se ha relacionado con la precisión con que se realiza una medida ([5]; [6, p. 11-12]). Para hacer clara la distinción entre generalidad y vaguedad, consideremos un conjunto de objetos que presentan la misma cantidad de atributo y otro conjunto de objetos que presentan diferente cantidad de atributo. Dentro de cada conjunto, todos los objetos presentan la misma cantidad de atributo, de tal manera que todos presentarían la misma vaguedad ya que con la misma propiedad se les podría aplicar un término para reflejar tal cantidad de atributo; sin embargo, a la hora de medir la cantidad atributo de los objetos incluidos en dicho conjunto (presuponiendo que todos los objetos lo presentan en igual cantidad), las variaciones en los valores atribuidos a dichos objetos, serían debidas a la falta de precisión del instrumento de medida y no al hecho de que tal atributo fuera vago. Sin embargo, cuando comparamos dos objetos pertenecientes a dos conjuntos diferentes, y por tanto presentando diferente cantidad de atributo, la vaguedad aparecería en cuanto hasta que punto es más licito aplicar el término que representa el atributo a un objeto frente a otro. Por tanto, la vaguedad aparece en cuanto un mismo término es utilizado para designar a objetos que presentan diferente cantidad de atributo, siendo más licito aplicarlo en unos casos que en otros. Sin embargo, el problema de la vaguedad se hace aún más patente cuando nos preguntamos por cual es el dominio de objetos a los que es extensible un término. Si suponemos un conjunto de referencia compuesto por todos los objetos posibles, habrá objetos para los que será muy adecuado aplicar el término y otros para los que será también claro no aplicarlo. Tales objetos coincidirán con aquellos que presentan mayor cantidad de atributo y aquellos que presentan menor cantidad de él, pero para los objetos que presenten cantidades intermedias de atributo, la vaguedad se podrá de manifiesto en cuanto hasta que punto es más lícito aplicar dicho término o dejar de aplicarlo.

Peirce ([7], [8]) considera que la vaguedad de un término procede del uso que los sujetos hacen de él. Esto ha llevado a considerar la consistencia con que se aplica un término a un conjunto de objetos, como una buena estimación de la vaguedad del mismo.

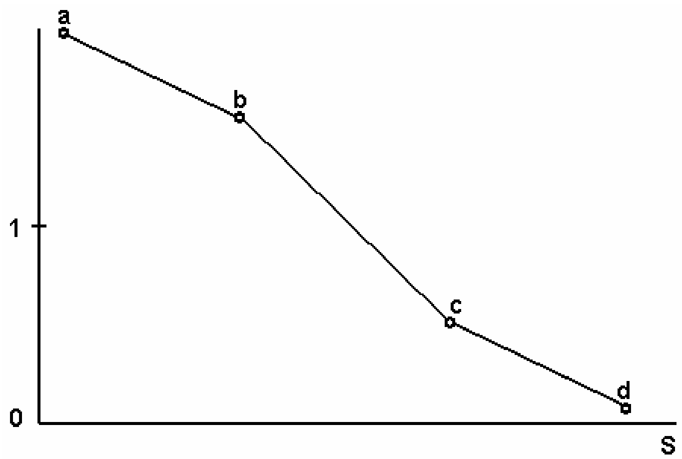
2.1 Black (1937)

Trata de cuantificar la definición de Peirce ([7], [8]), considerando que la consistencia es un buen índice de la vaguedad de un término. Se considera que habrá objetos para los que la aplicación de un término producirá pocas variaciones a lo largo de los sujetos que hacen uso de él, de tal manera que tal consistencia nos permitirá decidir si ese conjunto de objetos constituye el dominio de aplicabilidad del término o bien no.

Consideremos un conjunto de objetos S={s} y un término dado T. La consistencia (C) con que dicho término se aplica a cada uno de dichos objetos dentro de una población de sujetos, se puede estimar a partir de una muestra de estos, calculando la razón entre el número de sujetos que utilizan T para designar a s (M) y el número de aquellos que no lo hacen (N'), de tal manera que cuanto más se acerque el tamaño de esta muestra al de la población más cerca estará este valor del presente en dicha población:

$$C(T,s) = \lim_{\substack{M \to \infty \\ N' \to \infty}} \frac{M}{N'} \qquad (1)$$

Para un término T dado tendremos un valor C(T,s) para cada objeto s de S, de tal manera que si representamos gráficamente cada uno de los valores de consistencia que adoptan los diferentes objetos s, de forma hipotética tendremos una representación como sigue:



En este gráfico se pueden diferenciar tres zonas: (a-b), (b-c) y (c-d). La primera de ellas se puede considerar que constituye el dominio de aplicablidad de T, debido a que en esta zona todos los s de S tienen un valor C(T,s) claramente mayor que 1. De igual manera, la tercera zona constituye el dominio de no aplicabilidad de T. La zona intermedia es definida como la zona en la que el término presenta mayor ambigüedad, ya que se trata de la zona para la que los s de S adoptan valores de C(T,s) cercanos a 1 y en consecuencia existe aproximadamente el mismo número de sujetos que considera aplicable y no aplicable el término a dichos objetos.

Para hacer una estimación global del grado en que un término es vago, Black [5] considera que la pendiente de la recta que pasa por b y c es una buena estimación global del grado de vaguedad de dicho término. De esta manera, cuanto mayor es la pendiente menor es el grado de vaguedad de ese término, ya que la suma de los objetos incluidos en (a-b) y (c-d) tenderán a ser igual al número de objetos que componen S. En el caso de que la pendiente de la recta fuese infinita se tendría una división crispiana o nítida entre el conjunto de objetos s a los que son aplicables el término T y aquellos objetos a los que no es aplicable T (el segmento b-c sería perpendicular). En

esta situación también ocurriría que todos los objetos incluidos en estos conjuntos tendrían el mismo grado de vaguedad (los segmentos a-b y c-d serían horizontales).

Si en vez de un término, tenemos varios (T={T_i, ...,T_n}), éstos podrán ser caracterizados por los valores de consistencia que adoptan para cada objeto s de S. De este modo podremos hablar de T_i (s, c_i) y en consecuencia también de la consistencia con que se aplica a dicho objetos un término que sea de forma perfecta opuesto a T. Denominemos al opuesto de T como noT. Aquí hacemos la distinción entre la no aplicabilidad de T y la aplicabilidad de noT. Por ejemplo, no es lo mismo decir que una cosa no es roja a decir que esa cosa tiene el color opuesto al rojo.

Black [5] postula que dicha consistencia será la inversa de la consistencia con que se aplica T, es decir:

$$C(T,s) = \frac{1}{C(notT,s)} \tag{2}$$

De este modo para cualquier objeto s de S el producto de la consistencia con que se le aplica un termino T frente a la consistencia de su opuesto, será igual siempre a uno.

$$C(T,s).C(notT,s) = 1 \tag{3}$$

A la hora de contrastar empíricamente este modelo nos encontramos con varias dificultades:

- En primer lugar: la medida de consistencia no está acotada, de tal manera que ello dificulta la interpretabilidad de sus valores.
- En segundo lugar: para poder hablar en términos globales del grado de vaguedad de un término es necesario definir una métrica sobre los elementos de S de modo que puedan ser ordenados dentro del eje de abscisas, de no ser así la localización de los objetos en tal eje sería totalmente arbitraria y el calculo de dicha pendiente se haría imposible. Una segunda solución la constituye suponer una determinada relación entre la consistencia con que se aplica un término a un objeto y su posición en el eje de abscisas, de tal modo que se puede determinar tanto la consistencia global con que se aplica el término como la posición de cada uno de los objetos sobre S.

2.2 Hempel (1939)

Retoma el trabajo iniciado por Black [5] y realiza una serie de modificaciones en cuanto al cálculo de la consistencia con que se aplica un término.

$$C(T,s) = \lim_{\substack{M \to \infty \\ N' \to \infty}} \frac{M}{M + N'}$$

Con estas modificaciones consigue que los valores que adopta C(T, s) estén acotados, al estar comprendidos entre [0, 1]. Ahora el objeto s que presentará una mayor vaguedad será aquel para el que C(T, s) sea igual a 0,5.

Por otro lado, Hempel [9] trata de solucionar el problema de la necesidad de definir un orden en S, expresando el grado de consistencia global con que es aplicado un término, a partir de un índice del grado promedio en que los valores de consistencia que adoptan los objetos de S se alejan del valor 0,5. Hempel [9] habla de precisión con que se utiliza un término, sin embargo tal aspecto se relaciona con el concepto de generalidad, por lo que hablaremos de consistencia global como un promedio de la consistencia con que un término se aplica a los diferentes objetos de los que se compone el conjunto S, es decir, C(T,S).

$$C(T,S)=\frac{4}{N}\sum_{k=1}^{N}\left[C(T,s_k)-\frac{1}{2}\right]^2$$

Donde N representa el número de elementos presentes en S. Vemos que tal expresión es totalmente independiente de cualquier métrica definible en S.

La vaguedad del término vendrá dada por el complemento de la consistencia, es decir:

$$vg=1-C(T,S)$$

Black [5] afirmaba que la consistencia con que se aplicaba un término a un objeto era inversamente proporcional a la consistencia con que se aplicaba su término opuesto en significado a ese mismo término. Vamos a ver si tal afirmación se puede mantener dentro de las expresiones de Hempel [9].

Si partimos de que M representa el número de sujetos que en una población aplican el término T a un objeto s y de que N' es el número de sujetos que no aplican ese término al objeto s, podemos identificar C(T, s) con M y C(noT, s) con N'. De esta forma, la suma de M y N' constituyen el número de sujetos presentes en dicha población. Para hacer tal suposición es necesario considerar que la respuesta del sujeto es dicotómica, es decir, o bien dice que T es aplicable a s o bien dice que no es aplicable. Desde el punto de vista que estamos tratando el tema resulta razonable hacer tal suposición, al estar hablando en términos de si el sujeto aplica o no aplica ese término. Por lo tanto, vamos a considerarlo como un acto que se da en sí y no vamos a entrar en analizar como se podría medir, en cuyo caso, tal suposición, dependiendo de la tarea a emplear, podría cuestionarse.

Si utilizamos la expresión de Black [5], tendremos que:

$$C(T,s)=\frac{M}{N'} \qquad C(noT,s)=\frac{N'}{M} \qquad C(T,s)=\frac{1}{C(noT,s)}$$

En el caso de Hempel (1939) nos encontramos con la siguiente situación:

$$C(T,s)=\frac{M}{M+N'} \qquad\qquad C(noT,s)=\frac{N'}{M+N'}$$

$$C(T,s)+C(noT,s)=\frac{M+N'}{M+N'}=1$$

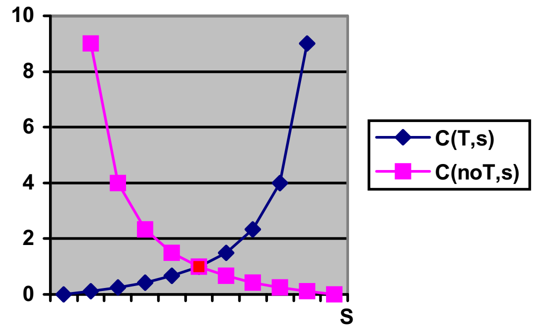
$$C(T,s)=1-C(noT,s)$$

En lo que se refiere a la consistencia global con que se aplica el opuesto de un término, partiendo de la expresión anterior podemos expresarlo del siguiente modo:

$$C(noT,S) = \sum_{k=1}^{N}\left[C(noT,sk) - \frac{1}{2}\right]^2 = \sum_{k=1}^{N}\left[1 - C(T,sk) - \frac{1}{2}\right]^2 =$$

$$= \sum_{k=1}^{N}\left[\frac{1}{2} - C(T,sk)\right]^2 = C(T,S)$$

Como consecuencia obtenemos que la consistencia global con que se aplica un término a un conjunto de objetos, coincide con la consistencia con que se aplica el término opuesto de este a ese mismo conjunto de objetos. En el caso de Black [5] la consistencia global con que T es aplicado a un objeto s es inversa a la consistencia con que noT se aplica a dicho objeto s, esto daría lugar a una representación gráfica en la que la curva de consistencia de T y noT se cortarían en un punto, siendo simétricas respecto a un eje vertical y presentando dos asíntotas verticales con limites en el infinito.



En la gráfica anterior se puede ver como la medida de consistencia aportada por Black [5] consiste en una medida del dominio de aplicabilidad de un término T sobre un conjunto de objetos S. En cambio la medida de consitencia de Hempel [9] consiste en una medida del dominio de aplicabilidad tanto del término T como de su opuesto. Para transformar la medida de consitencia de Hempel en una medida de aplicabilidad de un término, es necesario definir una escala en S, con lo que nos volveríamos a enfrentar con el problema de Black [5]. Este problema fue eludido por Hempel expresando la consistencia global con que se aplica un término a partir de los valores de consistencia con que dicho término se aplica a cada uno de los objetos de S. De esta manera nuestro discurso se limita a decir que cuanto mayor es la consistencia con que dicho término se aplica a cada uno de los objetos de S, mayor será la consistencia global de ese término. Por tanto, dicha medida de consistencia va a depender de los objetos presentes en S. Para hacer una estimación global de estas magnitudes es necesario realizar un adecuado muestreo entre los objetos a los que se aplica dicho término, dentro del universo S. El hecho de que Black [5] presuponiera una escala definida sobre S, aludia a tal muestreo: la posición de cada objeto sobre S era su posición en relación al resto de objetos del universo S, lo cual le permitía hablar directamente del dominio en que es aplicable un términio dentro de un universo de objetos S. Por tanto, la apuesta de Hempel no elude la necesidad de definir una escala en S.

Una de las aplicaciones más importantes de los SE ha tenido lugar en el campo médico, donde éstos han sido utilizados para recoger, organizar, almacenar, poner al día y recuperar información médica de una forma eficiente y rápida, permitiendo aprender de la experiencia [2].

Dentro de los trabajos relacionados con el tema se pueden citar: INTERNIST-1/CADUCEUS, sistema de fácil uso sobre medicina interna. MYCIN, construido en Stanford que diagnostica enfermedades infecciosas de la sangre y receta los antibióticos apropiados y PUFF que diagnostica enfermedades pulmonares.
En la Universidad de Edimburgo se esta realizando una tesis de doctorado para evaluar condiciones de monitoreo en neonatos en una UCI. Dicho sistema detecta falsas alarmas causadas por diferentes factores y puede inferir el comportamiento de una variable [3].

Desde hace 2 años, la Fundación Cardiovascular de Colombia (FCV) ha ido desarrollando una herramienta de ayuda diagnóstica y terapéutica orientada a la reducción del error humano y a la mejora de procesos de atención en salud gracias a la toma oportuna de conductas para preservar la estabilidad clínica y predecir la complicación de niños atendidos en la UCI postoperatoria cardiovascular pediátrica. Dicha herramienta, podría en un futuro ser extendida a otras disciplinas y centros de atención de alta complejidad, con enormes beneficios tanto en el sector público como en el privado.

3 Medición de las funciones de pertenencia

La función de pertenencia tiene como objetivo generalizar el criterio de inclusión de un conjunto de objetos en una categoría [10]. Dicha generalización es de utilidad cuando se está tratando de representar propiedades en las que la relación entre las magnitudes que presentan unos objetos y las etiquetas que se utilizan para comunicar tales magnitudes, es una relación de vaguedad.

Siguiendo el trabajo de Norwich y Turksen [11], para medir una función de pertenencia debemos manejar los siguientes tipos de relaciones.

En primer lugar partimos de un conjunto S de objetos que plasman diferentes estados de una determinada propiedad A (por ejemplo, personas con diferentes alturas) Para referirnos a dichos estados utilizamos un conjunto $T = \{t_i\}$ de términos lingüísticos t, de manera que como existen diferentes objetos a los que se les hace corresponder una misma categoría definida a partir de dicho término, el tipo de relación definida será de vaguedad. Para reflejar dicha relación se construye un conjunto difuso, el cual se define a partir de una función de pertenencia μ, la cual sería una aplicación sobre dichos objetos que tiene como espacio imagen el subconjunto [0, 1] de los números reales.

Cuando se dispone de un referente físico, como en el caso del atributo A altura, sería posible definir una función θ: $S \rightarrow \Re$ que asignase a cada persona u objeto s de S su altura con un valor real $\Re$, de tal manera que ahora la función de pertenencia μ para la etiqueta lingüística t de T (por ejemplo, muy alto), sería una aplicación μ_t: $\Re \rightarrow [0, 1]$, que refleja el grado en que cada valor numérico de la recta real refleja ade-

cuadamente el significado de la etiqueta t o a la inversa, el grado en que la etiqueta t es aplicable a dicho valor numérico.

Los principales procedimientos para obtener las funciones de pertenencia son los siguientes [12]:

1. **Selección de categorías**: se le presenta al sujeto un objeto y se le pide que seleccione la categoría a la que este pertenece, dentro de un conjunto de categorías posibles. Las categorías suelen venir ejemplificadas por los términos a los cuales se suele hacer referencia para designar a los objetos.
2. **Estimación de intervalos**: se le presenta al sujeto una etiqueta y se le pide que dé un intervalo de valores dentro de una escala numérica (por ejemplo de 0 a 100), de modo que se especifique el conjunto de valores para los cuales utiliza dicha etiqueta.
3. **Ejemplificación de la función de pertenencia**: al sujeto se le ofrece en el cuestionario una gráfica con dos ejes cartesianos. En el eje de abscisas se dispone una escala numérica (comprendida, normalmente entre 0 y 100), en la cual se supone que están dispuestos todos los estados que puede adoptar la propiedad que se está analizando. En el eje de ordenadas se dispone una escala numérica de valores comprendidos entre 0 y 1, en ella se disponen los valores de aplicabilidad de una categoría dada a cada uno de los estados diferenciados en el eje de abscisas. La gráfica se le dá al sujeto inicialmente en blanco y se le pide que dibuje la relación entre estas dos medidas para una etiqueta dada. Este dibujo suele ser una curva que refleja el grado en que cada uno de los estados de la propiedad está reflejado en el término. Dicho grado puede variar entre 0 y 1, siendo el valor que se señala en el componente segundo de las coordenadas bidimensionales de los puntos.
4. **A partir de datos estadísticos**: se parte de una distribución de frecuencias con las que diferentes valores numéricos se ven asociados con una etiqueta lingüística. Se normaliza tal distribución haciendo corresponder al valor con mayor frecuencia, el valor de pertenencia igual a 1, y al resto un valor proporcional a la distancia en frecuencias que guardaba respecto al anterior. Para la realización de esta normalización se han gastado muchos esfuerzos en desarrollar procedimientos de interpolación (por ejemplo, se puede consultar el trabajo de Chen y Otto [13]).
5. **Interfaz adaptable**: muy parecido a la Ejemplificación de la Función de Pertenencia, en este caso se le presenta por medio de un interfaz gráfico computerizado, una representación de la Función de Pertenencia construida para un término dado. El sujeto puede modificar la curva de pertenencia haciendo uso de los cursores del teclado del ordenador. Su tarea consistirá en modificar la curva de pertenencia de modo que con ello refleje el grado en que los valores del eje de abscisas están reflejados en el término. De esta forma irá definiendo los valores del eje de ordenadas, los cuales quedarán registrados en el ordenador.

El principal problema que presentan estos procedimientos es que solamente resultan útiles cuando se conoce la función θ que permite asignar valores numéricos a los objetos de *S*. Por ejemplo, cuando se pretenden estudiar propiedades como la gravedad de una serie de delitos o el grado en que unas frases expresan una postura autoritaria, la construcción de conjuntos difusos para reflejar el grado en que diferentes etiquetas son adecuadas para reflejar los diferentes grados de gravedad de un delito o

de autoritarismo de una frase, se ve limitada porque para estos objetos no se dispone de una función θ que presente unas propiedades métricas sólidas como una escala de intervalos o de razón. En muchas ocasiones, simplemente se pueden ordenar los delitos o los grados de autoritarismo que expresan las frases, siendo frecuente que los sujetos que se utilizan para construir θ comentan graves intransitividades en sus juicios., por lo que la problemática presente en la medición de las funciones de pertenencia se reduce a la existente en la medición de atributos psicológicos, y nuevamente llegamos a la misma conclusión que la vista en la medición de la vaguedad.

4 Conclusiones

Se ha realizado una revisión de las principales posturas existentes a la hora de diferenciar los conceptos de vaguedad, generalidad, ambigüedad y difuso. La vaguedad está vinculada al uso de términos lingüísticos para referirse a una propiedad continua, de modo que no se puede establecer un límite claro entre aquellos valores de la propiedad que son captados por el termino y aquellos que no. La ambigüedad está íntimamente relacionada con la ambigüedad en aquellos casos en los que para hacer referencia a la cantidad de atributo de un objeto, existen diferentes etiquetas que son igualmente aplicables. La generalidad, por su parte, hace referencia a una relación de equivalencia entre los objetos en referencia a una propiedad dada. Finalmente, el concepto de difuso puede ser asimilado al de vaguedad, donde la ambigüedad estaría vinculada a la existencia de objetos para los que no es claro si es aplicable o no una determinada propiedad o etiqueta lingüística que trata de reflejarla.

A la hora de construir conjuntos difusos para etiquetas lingüísticas que reflejan propiedades para las que no existe un referente físico, se plantea el problema de que los métodos existentes para medir las funciones de pertenencia, no son adecuados o más bien incompletos.

La construcción de conjuntos difusos para propiedades que no tienen un referente claro presentan una serie de dificultades que no han podido ser resultas de forma clara en Ciencias Sociales. Todavía se sigue discutiendo si es posible construir escalas numéricas para propiedades subjetivas. A pesar de que fuera posible resolver este problema, quedarían por resolver otros problemas asociados como ofrecer un modelo que permita reflejar las variaciones en el uso de un término lingüístico en función del contexto en el que se empleen. Así, se tiene constancia de que dependiendo del número de etiquetas lingüísticas que puede utilizar un sujeto y el tipo de tarea que realiza (por ejemplo, juicios absolutos frente a juicios relativos), varia la forma de la función de pertenencia ([14]).

Una posible solución a la problemática de la escalabilidad del referente puede venir de la eliminación del requerimiento de disponer un referente a la hora de modelar el uso que las personas hacen de los términos lingüísticos. En esta dirección, el enfoque de escalamiento basado en la medición conjunta (ver [15]) puede resultar útil ya que permitiría estudiar las preferencias de las personas a partir de cómo utilizan etiquetas lingüísticas que expresan cantidad, sin necesidad de que medie ningún tipo de representación numérica de las etiquetas. Sin embargo, esta alternativa nuevamente exige al sujeto que utilice las etiquetas lingüísticas con gran consistencia, no sólo no

convirtiendo intransitividades, sino satisfaciendo condiciones mucho más rigurosas como las denominadas condiciones de cancelación.

A pesar de que la definición de racionalidad de Dennet habla de la mejor elección a la luz del conocimiento disponible, la irracionalidad que parece gobernar los juicios que emiten las personas acerca de sus actitudes, solo puede ser atribuida a que no somos capaces de vislumbrar que tipos de representaciones pueden subyacer a esos juicios.

Finalmente indicar otra dificultad: las actitudes de las personas no parecen ser tan ricas como para poder ser escaladas utilizando valores numéricos. En la mayor parte de las ocasiones, el número de objetos, conductas o situaciones que las ponen de manifiesto es tan reducido en cuanto a número, que incluso los diseñadores de encuestas se ven en las dificultades para construir sus cuestionarios. Esta situación podría conducir a que se tuviera que optar definitivamente por utilizar el lenguaje natural, sin hacer referencia a ningún tipo de escala numérica, cuando se trata de estudiar científicamente las actitudes de las personas.

Referencias

[1] Haugeland, J. Mind Design II. Philosophy, Psychology, Artificial Intelligence. Cambridge: The MIT Press. (1997).

[2] Kelly, J. A. Entrenamiento de las habilidades sociales. Bilbao: Desclée De Brouwer. (2000).

[3] Hersh, H.M & Caramazza, A. A fuzzy set approach to modifiers and vagueness in natural lenguaje. Journal of Experimental Psychology: General, 105 (1976), 256-76.

[4] Zadeh, L.A. Fuzzy sets. Information and Control, 8 (1965), 338-53.

[5] Black, M., Vagueness: An exercise in logical analysis. Phylosophy of Science, 4 (1937), 427-55.

[6] Smithson, M. Fuzzy Sets Analysis for Behavioural and Social Sciences. New York: Springer-Verlag. (1987).

[7] Peirce, C. S. How to Make Our Ideas Clear, Popular Science Monthly 12 (1878), pp. 286-302

[8] Peirce, C. S. Vague, in J.M. Baldwin (Ed.), Dictionary of Philosophy and Psychology, 2 Volumes, London: Macmillan, p.748 (1902)

[9] Hempel, C.G. Vagueness and logic. Philosophy of Science, 6 (1939), 163-180.

[10] Lowen, R. Mathematics and fuzziness. En A. Jones, A. Kaufmann & H.K. Zimmermann (Eds.), Fuzzy Set Theory and Applications. Holland: D Reidel (1986).

[11] Norwich, A.M. y Turksen, I.B. A model for the measurement of membership and the consequences of its empirical implementation. Fuzzy Sets and Systems, 12 (1984), 1-25.

[12] Santamarina, C. y Salvendy, G. Fuzzy Sets based knowledge systems and knowledge elicitation. Behaviour and information technology, 10 (1991), 23-40.

[13] Chen, E.C. & Otto, K.V. Constructing membership functions using interpolation and measurement theory. Fuzzy Sets and Systems, 73 (1995), 313-27.

[14] Herencia, A.J. La vaguedad y el escalamiento de estímulos. Universidad de Granada, Facultad de Psicología: Departamento de Psicología Social y Metodología de las CC.CC. [Tesina no publicada]. (2001).

[15] Mitchell, J. An introduction to the logic of psychological measurement. New Jersey: Lawrence Erlbaum, (1990).

Bletchley Park: La emergencia de la computación según el modelo de cognición social distribuida

Alberto Rubio Frutos

Dto. de Lingüística, Lógica y Filosofía de la Ciencia. Facultad de Filosofía y Letras
(Universidad Autónoma de Madrid) Campus de Cantoblanco. Ctra. De Colmenar km. 15
(28049) Madrid
Alberto.rubio@uam.es

Resumen. Bletchley Park es conocido por ser el lugar donde durante la II Guerra Mundial, algunos de los mejores matemáticos británicos trabajaron para descifrar los mensajes en clave interceptados al Ejército Alemán. Para llevar a cabo su trabajo, se vieron obligados a construir diversos artefactos que contribuyeron a pensar que, por primera vez, se contaba con gran parte de los medios necesarios para el desarrollo de la Inteligencia Artificial. Sobre estos desarrollos tecnológicos se consolidaría posteriormente la fundamentación del "modelo clásico" en ciencias cognitivas. El sistema de trabajo que se desarrolló en Bletchley Park se puede describir bajo el modelo de "Cognición Social Distribuida" que, paradójicamente, supuso una puesta en cuestión del modelo tradicional cognitivo-computacional. Pretendo defender que un sistema de trabajo sobre las bases de la "Cognición Social Distribuida" fue el que consolidó las bases de las tareas que realiza un sistema computacional clásico, y que son precisamente las tareas que hoy podemos capturar bajo el modelo de "Cognición Social Distribuida", como los sistemas autónomos de navegación, las que conforman actualmente los mayores retos de la Inteligencia Artificial.

1 Introducción

El nacimiento de las ciencias cognitivas es una de las consecuencias más importantes del desarrollo de la Inteligencia Artificial durante el pasado siglo, ya que supone establecer un programa de investigación que involucra a distintas disciplinas, cuyas innovaciones en el seno de cada una de ellas suponen el enriquecimiento de su conjunto.

La emergencia de la computación tiene tres claras fases diferenciadas: en primer lugar, surge de un debate con importantes implicaciones filosóficas en relación a la fundamentación de la lógica y las matemáticas, en las pruebas de limitación de los sistemas formales por parte de Gödel, Church y Turing; en segundo lugar, el esfuerzo de los criptógrafos durante la II Guerra Mundial, para descubrir los mensajes secretos cifrados por los alemanes mediante la máquina Enigma, lo que conlleva importantes innovaciones en la tecnología necesaria para la construcción de los modelos teóricos diseñados en el seno del debate anterior; en tercer lugar, inspirados en la tecnología desarrollada durante la II Guerra Mundial, la construcción de las primeras computa-

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 12-21, 2006.

doras en el Reino Unido y en los EEUU significa la consolidación de un diseño computacional que ha pervivido prácticamente sin cambios hasta la actualidad, conocido como arquitectura “Von Neumann”.

La funcionalidad de estas máquinas abrió la posibilidad de que surgieran defensores del modelo computacional como un sistema que pudiera dar cuenta de las capacidades cognitivas humanas. De este modo, del debate sobre la fundamentación de las matemáticas y la lógica pasamos al debate en relación a la fundamentación de las ciencias humanas y sociales. El modelo cognitivo-computacional se instauró como paradigma dominante en las mismas. Sin embargo, a lo largo de las últimas décadas comenzaron a surgir una serie de dificultades en detrimento de la consolidación del programa. La “Teoría Representacional de la Mente”, consolidada sobre el modelo computacional, mantenía que el cerebro realizaba una serie de operaciones sintácticas sobre representaciones del mundo exterior, en una suerte de procesos en el interior del sistema que coincidían funcionalmente con los que se daban en el interior de una computadora. Las dificultades se concentraban en los excesivos costes que representaban el procesamiento en serie de las computadoras en relación a los sistemas neurales y en las fronteras demasiado marcadas entre el medio y el sistema cognitivo-computacional, que marcaban una pauta de falta de flexibilidad y autonomía de los modelos computacionales clásicos.

Sobre este punto se fueron consolidando modelos cognitivos diferentes, como los conexionistas, los Sistemas Dinámicos o las posturas que defendían una cognición extendida sobre el ambiente. Dentro de esta última corriente surgió dentro del campo de la antropología cognitiva de la mano de Edwin Hutchins, en la presentación de su modelo de “Cognición Social Distribuida”. Este modelo está basado en la denuncia por el escaso interés que ofrecían los modelos anteriores al desarrollo histórico y cultural de los procesos cognitivos, así como a los procesos de mediación entre subsistemas formados por seres humanos, artefactos y diferentes elementos del medio, integrados todos ellos en un sistema formalizado.

2 Bletchley Park

Bletchley Park comenzó siendo una casa de campo pocos meses antes del comienzo de la Segunda Guerra Mundial y en el desarrollo de la misma, pasó a extender su influencia no sólo a otros edificios de la misma localidad, sino incluso a otras ciudades. Su parentesco más cercano es el “Goverment Code and Cypher School” (GC&CS), un departamento del servicio postal británico, que a partir de la Primera Guerra Mundial comenzó a contratar criptólogos para interceptar los mensajes telegráficos alemanes que llegaban a sus costas. Cuando el GC&CS decidió centralizar su actividad en un sólo lugar y llegó a Bletchey Park, trabajaban en él alrededor de unas 100 personas Llegaron a siete mil en 1944 y casi a nueve mil al acabar la guerra.

Sus mayores éxitos se concentran en el diseño de Ultra, una máquina que fue fundamental en tres momentos (Agar, 2003): en primer lugar, al romper el código que utilizaba la armada italiana y el ejercito alemán en 1941, lo que supuso los primeros reveses de la campaña de Rommel en el Norte de África; en segundo lugar, al descifrar la Enigma de la armada naval alemana desde Junio del 41 a Enero del 42, y de

nuevo en Diciembre del 42, lo que supuso la hegemonía naval británica en el Atlántico Norte; por último, contribuyó de manera decisiva en el desembarco de Normandía, descifrando la clave "Fish"

Bletchley Park estaba separado en "barracas" militares. La barraca 6 se encargaba de decodificar los mensajes interceptados por las estaciones "Y", transmisiones de alta velocidad cuya Enigma era la "Roja", utilizada por la fuerza aérea alemana. Las claves se cambiaban cada día, y los criptólogos trabajaban a contrarreloj sobre los errores que cometían los operadores alemanes en el código Morse, con la ayuda del diseño de las "Bombes", máquinas diseñadas por los criptógrafos polacos con el objetivo de romper los mensajes de Enigma. Si el desciframiento del mensaje tenía éxito, entonces la información pasaba a la barraca 3, donde se juzgaba su importancia. Esa información se mandaba directamente a Londres.

El caso de la Enigma utilizada por la fuerza naval alemana era un reto todavía mayor. En un informe (Turing y otros, extraído de Teucher, 2004) escrito a finales de 1939, y firmado por los responsables la barraca 8, entre los que destaca el propio Turing (que había desarrollado el primer modelo computacional teórico en el seno del sobre la fundamentación de las matemáticas), se considera imposible descifrar los mensajes si no se construye lo que ellos denominaron "superbombe machine". El ataque a la Enigma naval era prioritario en Bletchley Park, por este motivo los mejores recursos y el mejor personal, incluido Alan Turing, fue destinado allí. Los escasos éxitos de la Barraca 8 eran destinados a la barraca 4, en los que una serie de distintos subdepartamentos se encargaban del mismo modo que en la barraca 3 de juzgar la prioridad del mensaje para después traducir la información al inglés.

En Octubre de 1941 la situación era desesperada. Distintos miembros de las barracas 6 y 8 que se sentían frustrados por la falta de medios, entre los que también, como no, se encontraba Turing, exigieron al propio Churchill más recursos. El primer ministro en persona les dio carta blanca. En ese momento es cuando Bletchley Park pasa a ser una industria de producción de información, y no un grupo de descifradores de códigos formados por los más destacados jóvenes matemáticos e ingenieros del Reino Unido. La evolución se puede dividir en tres etapas:

1. El trabajo individual de desciframiento con un lápiz, papeles y la ayuda de la secretaria.
2. La división estricta del trabajo perfectamente reglado por motivos de eficacia y seguridad, que incluía el diseño y la utilización de artefactos como "Bombes".
3. La producción industrial de información mediante la utilización de la más alta tecnología de la época, que a su vez dividimos en dos partes:

   3.1. El diseño de Ultra, basado en la máquina de Hollerith y en el viejo sistema de tarjetas perforadas
   3.2. La construcción a finales de 1943 de Colossus, protagonizada por Max Newman.

Según este mismo proceso pasamos de las habilidades de un cerebro con la ayuda de un lápiz, un papel y una secretaria, a un sistema complejo de personas y máquinas, a un monstruo computacional. La apelación a las máquinas se hizo necesaria porque

el desarrollo de ese trabajo por parte de los hombres era inviable. Sin embargo, lo que parece que se suele olvidar de esta historia es que hombres y mujeres eran los protagonistas tanto del diseño y el desarrollo de las máquinas. Uno de los resultados finales de ese trabajo fue la construcción de "Colossus", a través del diseño y construcción de las distintas "Bombes" y "Ultras". "Colossus" hacía parte del trabajo que le correspondía a las barracas dedicadas al desciframiento de los distintos Enigmas, que no sólo incluía el de la fuerza aérea y naval, sino que también por el ejército, los ferrocarriles y por el propio Alto Mando alemán.

La división del trabajo se establecía por motivos de eficacia, ya que la producción de información era descomunal, sino que es también importante considerar el celo con el que se conservaban las medidas de seguridad. Los miembros de una barraca no tenían acceso al trabajo de los de otras, ya que no se permitía que compartir la información de las otras, para que nadie tuviera una posición privilegiada y así dificultar el trabajo de posibles agentes dobles para Alemania. Dado el número ingente de personas que trabajaban en Bletchley Park estas medidas formaban parte de un protocolo de seguridad indispensable. La información que se trasladaba de una barraca a otras era la de apoyo al trabajo de éstas, pero se mantenía una estricta separación entre los medios y el contenido de la información sobre la que trabajaban. De hecho, el resultado global del trabajo en Bletchley Park sólo se obtenía en Londres.

Podemos establecer una fácil analogía entre las prácticas de Bletchley Park y las de un buque de guerra de la Armada Británica. De hecho, a medida que la guerra avanzaba, Bletchley Park iba adaptándose a la estructura de trabajo industrial a una estructura militar. Independientemente de las semejanzas y diferencias entre el espacio de producción industrial y el cuartelístico, a medida que el trabajo en Bletchley Park se iba haciendo decisivo en el desarrollo de la guerra, las condiciones de disciplina, vigilancia y seguridad, se iban acentuando. Esto significa que las prácticas de recepción, elaboración y producción de información, se regían según protocolos militares, con la consiguiente división estricta de la responsabilidad de las distintas barracas y sus componentes.

Por ejemplo, el proceso de navegación de un buque británico no exigía que cada uno de sus componentes tuviera que tener toda la información relativa a los objetivos que desde el Alto Mando se imponían a ese buque. Más bien se trataba de que cada uno hiciera bien su trabajo. Sin embargo, el resultado global es que el Alto Mando británico intercepta un mensaje enviado a Rommel o que un buque alemán es sorprendido y destruido en el Atlántico Norte. Es posible entonces analizar el trabajo desde la perspectiva global de la unidad de los distintos sistemas controlados desde Londres.

3 De la Criptología a la Inteligencia Artificial

Los matemáticos británicos comprendieron pronto, en 1939, que la única respuesta viable a la manipulación mecánica de información era la construcción de artefactos que realizaran las mismas tareas con mayor potencia y de forma inversa. Las máquinas "Ultra" y "Colossus" procedían realizando cálculos estadísticos sobre la frecuencia de aparición de las letras del alfabeto alemán. El proceso de desciframiento estaba

basado en la manipulación física, a través de mecanismos electrónicos, de símbolos que se correspondían a dichas letras.

Sin embargo, la tarea que realizaban estas máquinas no podía ser considerada como "inteligente" ya que el conjunto de algoritmos que ejecutaban estaban delimitados por los procedimientos específicos que eran necesarios para llevar a cabo el desciframiento. Para que una máquina fuera considerada como "inteligente" tenía que poder llevar a cabo un rango ilimitado de tareas, o lo que es lo mismo, debería ser capaz de calcular todas las funciones computables.

La tecnología que se había desarrollado para el diseño de "Ultra" y "Colossus" hacía posible la construcción de este artefacto, y la posibilidad teórica ya había sido desarrollada años antes por Alan Turing con su "Máquina Universal". Su diseño estaba contextualizado precisamente dentro de las limitaciones de los sistemas formales, en concreto, en la demostración de la imposibilidad de resolver el "Problema de la Decisión". Por este motivo, el propio Turing estuvo interesado el resto de su vida en desarrollar modelos teóricos alternativos al computacionalismo clásico del que fue pionero, entre los que cabe destacar su interés por la "morfogénesis" y sus incursiones en el "conexionismo" a finales de los años cuarenta. Sin embargo, a pesar de este escepticismo, Turing se hizo cargo de la posibilidad de la Inteligencia Artificial, tras comprobar la potencia de los artefactos que había contribuido a diseñar en Bletchley Park.

"Ultra" era capaz de realizar una tarea que un ingente grupo de personas bien organizada y preparadas no era capaz de realizar. Del mismo modo, una persona podía realizar un número ingente de tareas inteligentes que "Ultra" era incapaz de llevar a cabo. La posibilidad de la "Inteligencia Artificial" pasaba entonces por convertir un artefacto que llevaba a cabo una tarea específica de procesamiento de información, a convertirlo en un procesador universal. Y este salto de un modelo a otro, era el mismo que se dio al pasar del diseño de una "Máquina de Turing" a una "Máquina Universal".

Hay una gran confusión histórica en relación a la naturaleza de la "Máquina de Turing". En principio, podríamos decir que se trata de una "Máquina de Turing" consta de dos partes: un lector-borrador-marcador con una serie de instrucciones, que se mueve sobre una cinta separada por celdas en las que hay distintos símbolos. El lector modifica el contenido de la cinta en virtud de sus instrucciones y de este modo se lleva a cabo la computación de funciones. Una "Máquina Universal de Turing" es una "Máquina de Turing" con una configuración tal que podría simular el funcionamiento de cualquier otra "Máquina de Turing", cuya información está codificada en la cinta. De este modo, se podría llegar a construir una máquina que fuera capaz de realizar las mismas tareas que llevamos a cabo los seres humanos, incluso las criptológicas, mediante el diseño de la configuración de la máquina, el "hardware", y los distintos programas que podría llevar a cabo, el "software".

4 El funcionalismo computacional

La primera computadora con una estructura funcional similar a las actuales fue el EDVAC, desarrollada en los EEUU por un grupo de investigadores encabezados por John Von Neumann. Aunque el protagonismo de este último sea discutible, y posiblemente se construyera en el Reino Unido una computadora similar con anterioridad, la estructura funcional de nuestras computadoras se conoce como arquitectura "Von Neumann", y está compuesta principalmente por dispositivos de "output" e "input", un sistema de memoria, y un sistema de control central.

Los nuevos artefactos contaban con una serie de modificaciones en relación a su antecesor, el ENIAC, y al diseño de la "Máquina Universal de Turing", que se concentran en la relación entre un amplio sistema de memoria y el sistema de control central. Este nuevo diseño favorecía una mayor autonomía de la máquina en relación a los investigadores, lo que suponía una mejora en la eficacia del sistema. Lo que condujo a su vez, al fortalecimiento de la asunción de que los procesos intelectuales se daban en el interior de los sistemas, y de que la mejor manera de interpretar una máquina de Turing era mediante la implementación de la misma en un sistema "Von Neumann".

En la década siguiente, tras el desarrollo de las primeras computadoras, Newell, Shaw y Simon, llevan a cabo el proyecto de la arquitectura SOAR, en el contexto de la construcción del diseño de sistemas que pudieran dar cuenta de un conjunto muy amplio de problemas que iban desde el diseño de algoritmos hasta las dificultades que tenemos las personas en la conciliación de la vida laboral y profesional. El proyecto, a pesar de que en principio fue fruto de las intuiciones de Newell y Simon sobre las posibilidades de aplicación de los nuevos diseños computacionales, tiene su punto de referencia histórico en la defensa que hace Turing (1950) de la posibilidad de la Inteligencia Artificial sobre la base de lo que posteriormente se denominó "Test de Turing". Dicho test trata de mostrar que si el comportamiento de una máquina no puede ser discriminado del de una persona, por parte de un juez competente, en este caso, por parte de otra persona, nos vemos obligados a señalar que dicho artefacto demuestra inteligencia.

Los nuevos diseños computacionales tuvieron pronto el interés de filósofos y científicos sociales. Si estos diseños llevan a cabo los mismos procesos y obtienen los mismos resultados que los agentes humanos, entonces podríamos llegar a admitir que la simulación de estas tareas puede ser un marco privilegiado para el estudio de la mente humana. De este modo se consolida un determinado modo de pensar en la mente humana mediante el uso de la metáfora computacional, lo que se ha dado en llamar "Cognitivismo Clásico", "Funcionalismo Computacional", o también conocido como "Modelo Clásico" en Ciencia Cognitiva, que parte de la premisa que los procesos cognitivos que llevamos a cabo los seres humanos, son manipulaciones sintácticas de representaciones externas, esto es, son equivalentes a los procesos de manipulación símbolica que se dan en los sistemas computacionales tradicionales (Pylyshyn, 1984).

5 Cognición social distribuida

A la hora de emplear la metáfora computacional, uno de los mayores problemas que trató de solventar la ciencia de lo mental, se encontraba en la conciliación de los aspectos normativos de la ciencia y del sentido común de la psicología popular. Por ejemplo, si alguno de ustedes quiere prever donde voy a estar el domingo por la tarde, la mejor manera de averiguarlo no es realizar un estudio de mis movimientos a lo largo de la semana sino, más bien, preguntándomelo.

Si volvemos a la máquina “Ultra”, lo que ésta hacía era “preguntar” a “Enigma” el lugar exacto dónde se iban a encontrar los buques de guerra alemanes en el Atlántico Norte. Los buques británicos tenían que llegar a esta información a partir de los movimientos de sus adversarios. En ambos casos, se hacían complejos cálculos para responder a la pregunta, estaban involucrados un gran conjunto organizado de personas y artefactos, y había una larga tradición cultural que respaldaba ambas prácticas. ¿Cuál es la diferencia esencial? Pues que hemos encontrado instrumentos eficaces para resolver la cuestión de “Enigma”, pero todavía los artefactos diseñados hasta la fecha son ineficaces para resolver la segunda.

De la misma manera, el “Cognitivismo Clásico” desarrolló muy buenos modelos para dar cuenta de las conductas lingüísticas de los organismos, pero no así de su “ser en el mundo”. Mediante el uso de la metáfora computacional se podía establecer fácilmente que lo que se daba en el interior del cráneo eran una serie de procesos sintácticos, que manipulaban información semántica a través de representaciones del mundo, para que esto fuera posible debía haber una suerte de “lenguaje del pensamiento”, formado por una serie de subestructuras lingüísticas independientes de la sintaxis de los lenguajes naturales. Pero era muy difícil diseñar artefactos que se movieran con solvencia a través de una habitación, o que pudieran coordinar los movimientos de su propia estructura corporal.

La emergencia de la computación se ubica en un momento en el que era más efectivo aceptar el reto de “preguntar” a la máquina “Enigma” donde estaban los buques alemanes, frente a la posibilidad de llevar a cabo diseños destinados a mejorar a eficacia de la situación de los sistemas en el ambiente, como podían ser los buques de los “aliados” en alta mar.

De la mano de la antropología cognitiva, Edwin Hutchins (1995), trató precisamente de buscar cuáles eran las prácticas sociales y cognitivas que se daban a la hora de manejar un buque de la Armada Norteamericana. En su condición de antropólogo, una de sus más sorprendentes conclusiones, fue la de la denuncia de la presunción por parte del “cognitivismo clásico” de que las tareas inteligentes se daban en el interior de los sistemas. Hutchins, en primer lugar, compara las prácticas de navegación occidental con las empleadas por los aborígenes de las antípodas, para dar cuenta de que los procesos cognitivos que llevamos a cabo tienen una fundamental carga histórica; a continuación, dibuja un esquema de funcionamiento de un buque sobre un modelo que denomina “Cognición Social Distribuida”, en el que hay un protagonismo compartido entre los subsistemas que conforman artefactos y personas, y las contingencias del ambiente. Sobre este sistema imperan las prácticas de negociación entre los distintos agentes humanos, la manipulación de información a través del medio físico y de los artefactos, y no exclusivamente a través de los informes verbales.

El modelo de la "Cognición Social Distribuida" parte de la premisa de que el conjunto formado por el ambiente, los artefactos y los agentes humanos, conforman en sí mismos el sistema cognitivo, y no son un grupo de sistemas coordinados con diferencias efectivas entre el ambiente y el barco, entre los distintos agentes humanos, y la relación de estos últimos con los artefactos de navegación. Estos agentes conforman subsistemas de un sistema de orden más general, en el que las fronteras entre lo humano y lo ambiental, entre lo natural y lo artificial quedan desdibujadas. Es decir, la cognición queda "distribuida" o "extendida", sobre agentes de orden social y natural. Si apelamos exclusivamente al funcionamiento de las neuronas, cerebro por cerebro, de todos los agentes humanos implicados en el sistema, nunca lograremos adivinar cuál es el rumbo del barco. Debemos entonces comprender la dinámica del conjunto de subsistemas, para entender el funcionamiento del sistema cognitivo que conforma la embarcación y el medio ambiente sobre el que navega.

Si volvemos a Bletchley Park descubrimos que allí se dio el primer paso para el desarrollo físico de los computadores, en efecto, en un sistema de "Cognición Social Distribuida", en el que se seguían las prácticas de negociación, de interacción con artefactos, de manipulación de información, que recibía "inputs" y emitía "outputs", donde cada una de las "barracas" tenía que hacerse cargo de sus propias responsabilidades, todo ello con el objetivo común de descubrir cuáles iban a ser los movimientos de un sistema físico ajeno y provisto de intencionalidad.

6 Nuevas perspectivas

Los sistemas multi-agente representan en la actualidad el núcleo de los estudios que pretenden abordar desde una perspectiva distribuida y extendida, la posibilidad de llevar a cabo modelos computacionales que permitan integrar la ejecución de tareas de manera más eficaz, mediante la construcción de diversos agentes que realizan diversas sub-tareas para la concreción de un resultado final. Sin embargo, el mayor reto para que un sistema pueda adecuarse a las características de un modelo de "Cognición Social Distribuida" no es el coordinar distintas tareas por parte de diversos agentes, sino buscar la integración del ambiente real como un agente genuino del proceso. En muchos modelos multi-agente el ambiente se plantea adecuadamente como uno de los agentes del proceso. Sin embargo, esto supone a su vez que el ambiente sea excesivamente un constructo artificial, y resulte menos interesante. Por ejemplo, si quisiéramos desarrollar un gigantesco robot que surcara los mares las mayores dificultades estribarían en determinar de qué manera el agente ambiental externo variable e inesperado como es el océano, pudiera contenerse como uno de los sub-sistemas que ejecuta diversas tareas de manera coherente al resto.

En esta dirección, el éxito de los sistemas multi-agende depende de otro programa de investigación encabezado por Brooks (1986) que ya desde los años ochenta, comenzó a plantearse la eficacia de los sistemas representacionales, que imponían una frontera estricta entre el agente computacional y el medio. Entonces pensó en la alternativa de considerar el propio mundo como el mejor modelo. En la misma dirección, pero desde perspectivas algo más radicales, los Sistemas Dinámicos buscan nuevos modelos formalizados de manipulación de la información basados en disciplinas

diferentes a las matemáticas o la lógica, como es el caso de la Física.

Uno de los ejemplos más interesantes de laboratorios, esforzados en la conciliación de los aspectos más clásicos de la computación con los más novedosos, es el "Embodied Intelligence Laboratory" de la Universidad del Estado de Michigan (EEUU), encabezado por Juyang Weng. Sus esfuerzos se concentran en crear robots con capacidad de desarrollo, y para ello articula en el interior del mismo una arquitectura clásica "Von Neumann" conectada a su vez sobre un esquema independiente de mayor nivel, que tendría que vérselas con la variabilidad del ambiente.

Este tipo de esfuerzos van encaminados hacia el desarrollo de modelos integrados que vinculen de manera efectiva los procesos cognitivos de manipulación sintáctica de representaciones, (es decir, el cometido de "Ultra") con aspectos más dinámicos que ahora sólo podrían ser desarrollados por sistemas de "Cognición Social Distribuida", de la misma manera que eran estos los que se encargaban de llevar a cabo las tareas que ahora ejecutan fácilmente nuestros ordenadores personales.

7 Conclusiones

Desde el nacimiento de la computación clásica se ha tenido presente la posibilidad de que éste no fuera el único modelo posible, para dar cuenta de los procesos cognitivos humanos. He intentado esbozar a lo largo de esta exposición, los motivos históricos que llevaron a una determinada perspectiva sobre la naturaleza de la cognición. Dicha perspectiva se gestó sobre la resolución de una serie de problemas, que parecían coincidir con aspectos esenciales de la inteligencia humana, mediante el diseño de artefactos electrónicos.

El reto de tener que realizar tareas que no pueden llevar a cabo los seres humanos por su magnitud, pero que serían capaces de realizar si fueran de una menor escala, como la realización de pequeños cálculos o el dirigir una pequeña embarcación, se sitúa en el centro de la innovación tecnológica en Inteligencia Artificial desde dos distintos frentes. Las tareas se llevan a cabo primero por un modelo de "Cognición Social Distribuida" y cuando ésta no da a basto, se construyen artefactos que puedan dar lugar a la realización de las mismas. En el primer caso, los modelos teóricos y aplicados de computación clásica se fundieron por primera vez en Bletchley Park, lo que condujo a una determinada teoría de la cognición. En el segundo caso, las limitaciones impuestas por los sistemas representacionales con el protagonismo de sistemas de manipulación sintáctica, a la hora de mostrar la flexibilidad y la autonomía necesarias para dar cuenta de la variabilidad de las contingencias del ambiente, conducen a un renovado interés por modelos "externistas".

A lo largo de los próximos años, seremos testigos de una revolución tecnológica en esta dirección, no ya sólo de aspectos netamente computacionales, sino que se integraran esfuerzos de distintas disciplinas, en un intento por automatizar tareas hasta el momento sólo propias de los animales, en un intento de, como decía Andy Clark, poder llegar a construir un coche con el cerebro de una cucaracha.

Referencias

1. Agar, J. (2003) The Goverment Machine. A Revolutionary History of the Computer. Cambridge, MA: MIT Press
2. Brooks, R. A. (1986), A robust layered control system for a mobile robot, IEEE J. Robotics and Automation 2(1), 14-23
3. Clark, A. (1997) Being There. Putting Brain, Body and World Together Again, Cambridge, MA: MIT Press
4. Hodges (1983) Alan Turing: The Enigma. New York: Simon & Schuster
5. Hutchins E. (1995) Cognition in the Wild. Cambridge, MA: MIT Press
6. Pylyshyn, Z. (1984) Computation and cognition: Toward a Foundation for Cognitive Science. Cambridge, MA: MIT Press
7. Teucher, C. (2004) Alan Turing: Life and Legacy of a Great Thinker. Berlin: Springer-Verlag
8. Turing, A. (1936) On computable numbers with an Application to the Entsceidungproblem. Proceedings of the London Mathematical Society, ser. 2, vol. 42, 230-265.
9. Turing, A. (1950) Computing machinery and intelligence, Mind, 59 (236): 433-460
10. Weng, J. (2004) Developmental Robotics: Theory and experiments, International Journal of Humanoid Robotics, Vol. 1 No. 2 1999-236

Sobre la frontera formal entre el conocimiento computable y el conocimiento humano

Juan Carlos Herrero[1], José Mira[1], María Taboada[2] y Julio Des[3]

[1] Departamento de Inteligencia Artificial, E.T.S.I. Informática, UNED,
28040 - Madrid, España
jmira@dia.uned.es
[2] Dpto. de Electrónica e Computación, Universidad de Santiago de Compostela.
15782 Santiago de Compostela, España
chus@dec.usc.es
[3] Servicio de Oftalmología, Hospital Comarcal Dr. Julián García.
27400 Monforte de Lemos, España
eljjdes@telefonica.net

Resumen. Una parte importante del trabajo en Ingeniería del Conocimiento (IC) consiste en (1) modelar descripciones en lenguaje natural de una tarea y un método para resolverla computacionalmente y (2) encontrar procedimientos sistemáticos y automatizables de reescritura formal de esos modelos. En este trabajo mencionamos la situación actual de estas dos fronteras de la IC, proponemos un procedimiento de enlace entre el lenguaje natural y los lenguajes de programación, a través de un modelo estructural compartido, e ilustramos su aplicación a la tarea de diagnóstico en el dominio de la oftalmología. La posibilidad de generar código ejecutable a partir de un modelo a nivel de conocimiento para un conjunto de tareas y métodos cada vez más amplio es un objetivo claro y preciso de la IC. Adicionalmente, si se preservan las tablas de cambio de semántica en la reescritura formal, quedará claro en cada aplicación cuál es la parte del conocimiento que finalmente reside en la máquina y cuál la que permanece en el dominio del observador externo, a la espera de ser usada para documentar los resultados del cálculo.

Palabras clave: modelado del conocimiento, tareas genéricas, ontologías, generación automática de código ejecutable, implementación, diagnosis en medicina, grafos, árboles.

1 Introducción

El objetivo global de la Inteligencia Artificial (IA) es contribuir a hacer computable la mayor parte posible del conocimiento humano sobre tareas cognitivas y científico-técnicas. Para alcanzar esta meta, y siempre que el conocimiento al que hacemos referencia no sea situado (no necesite un cuerpo que lo soporte), usamos una serie de organizaciones superpuestas que facilitan la reescritura de nuestros modelos desde el lenguaje natural hasta el lenguaje máquina. Una parte importante del problema de la IA está entonces en especificar: (1) La frontera del lenguaje natural con los procesos cognitivos a los que pretende describir y (2) la frontera del primer lenguaje formal (el

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 22-33, 2006.

más próximo al natural) para el que ya existe un programa traductor capaz de llevarla a la máquina sin pérdida ostensible de semántica.

Una vez que tengamos claras estas dos fronteras también se hacen más claras las tres subtareas básicas de la perspectiva aplicada de la IA:

1. Conseguir incrementar la riqueza, expresividad y precisión de nuestras descripciones en lenguaje natural de los procesos cognitivos.
2. Desarrollar entornos de programación cada vez más próximos al lenguaje natural.
3. Desarrollar procedimientos sistemáticos y eficientes para enlazar (1) con (2).

Es decir, para reescribir formalmente las entidades del modelo conceptual, dejando una traza clara y explícita (una tabla de correspondencias en extenso) de los cambios de semántica y causalidad que se producen al pasar del lenguaje natural al formal.

En este trabajo no vamos a abordar el problema de la frontera del lenguaje natural con la cognición. Nos vamos a centrar en los otros dos. Es decir, vamos a comentar la evolución de los entornos de programación y vamos a proponer un procedimiento efectivo y sistemático de reescritura de modelos conceptuales en términos de lenguajes de programación. El procedimiento propuesto es válido para las tareas de análisis [1].

Partimos de la conjetura de que existe una conexión causal entre los mecanismos neuronales de un experto, un médico por ejemplo, y la descripción verbal o escrita que ese mismo experto hace sobre su proceso de razonamiento en determinadas tareas, por ejemplo en el diagnóstico. Esta conjetura subyace a todo el trabajo desarrollado en el paradigma representacional o simbólico de la Ingeniería del Conocimiento (IC), y también en gran parte de las aproximaciones situada y conexionista.

También aceptamos la conjetura de que es posible relacionar la sintaxis con la semántica de forma que una parte relevante del conocimiento soporte de un texto escrito se encuentra en las entidades y relaciones asociadas a las palabras o grupos de palabras que constituyen ese texto.

Estas dos conjeturas han guiado la mayoría de los desarrollos de entornos para facilitar la tarea de modelar (elicitar) el conocimiento, tales como CommonKADS [2,3,4], Protégé [5], algunos de los desarrollos de nuestro grupo [6,7,8] y ciertos entornos comerciales como el de TIBCO [9], un entorno de "integración de aplicaciones" (EAI) cuyo entorno gráfico de modelado (llamado Designer) integra el modelado de ontologías a través de XML y contiene un embrión de tareas genéricas de bajo nivel que puede ser inmediatamente ampliado a conjuntos de tareas genéricas en el sentido KADS, además de poderse basar también en un estándar como es XSLT (junto a XML).

Protégé es el primer entorno de desarrollo de aplicaciones en el que se contempla un modo genérico y reutilizable de modelado de ontologías (por medio de marcos, o siguiendo el modelo clase-objeto), además de ser gráfico, lo que le dota de gran potencia y ergonomía, e incorporar estándares como el UMLS. En este entorno, la explotación del conocimiento se puede hacer por medio de sentencias SQL (queries). Dicho conocimiento puede ser utilizado por desarrollos específicos, pero Protégé no contiene herramientas para modelar e implementar tareas genéricas con las que las ontologías puedan acoplarse.

2 Planteamiento del problema

Consideremos el siguiente supuesto: un observador C trata de resolver computacionalmente el problema A, que habitualmente es resuelto por un humano experto. Esto supone que estamos en realidad ante dos problemas de índole diferente:

1. Resolver el problema A
2. Resolver el problema B, que consiste en obtener la aplicación software que resuelve el problema A.

La naturaleza del problema A es diferente de la naturaleza del problema B. No olvidemos que la resolución del problema A se lleva a cabo por un experto humano, de hecho, independientemente de que existan computadoras que pudieran resolverlo; por tanto, la resolución del problema B no es necesaria para resolver el problema A, sino para resolver el problema A computacionalmente. Dicho de otro modo, la afirmación de que A puede resolverse computacionalmente debe ir acompañada de la necesaria justificación, que debe ser proporcionada mediante la resolución del problema B y una adecuada teoría que lo sustente, donde tal justificación ha de aparecer explícitamente. Así pues, para que el observador C (con quien nos podemos identificar cada uno de nosotros) obtenga su sistema de solución automática del problema A, parte de su trabajo va a ser de índole científica y parte será ingeniería.

La parte científica, o de análisis, consiste en la elaboración del modelo de conocimiento del problema A, que puede abarcar aspectos diversos en cada caso, tales como memoria, razonamiento, aprendizaje, interpretación del lenguaje, etc, implicados en la resolución del problema A pero que no están, en general, explícitos en la teoría de A, ni tienen que ver con el problema A, sino que son descubiertos, deducidos o ideados por el observador C. Así pues, contamos con la inteligencia natural para llevar a cabo esta parte.

La parte de ingeniería, o síntesis, consiste en la obtención de una aplicación software partiendo del modelo de conocimiento. A dicha síntesis la llamaremos síntesis simbólica, puesto que el resultado es una aplicación software que llamaremos modelo simbólico. Veremos que la síntesis simbólica, aunque puede ser llevada a cabo por el observador, puede ser automatizada.

El camino metodológico propuesto [1,7,13,16] considera que, una vez planteado un modelo de conocimiento, es posible obtener un modelo simbólico (y por tanto, computable) a través de la identificación de estructuras abstractas comunes al modelo de conocimiento y al modelo simbólico. Si afirmamos que un problema A, descrito a través de un modelo de conocimiento, puede ser reproducido por medio de una computación, esto equivale a afirmar que consideramos que el modelo de conocimiento en sí mismo es computable también o, dicho de otro modo, que debe haber algo que se entiende por computación, independientemente de que hablemos de un modelo de conocimiento o de un modelo simbólico. Dicho en términos de la teoría de niveles de descripción de un cálculo [10,11], debe haber un modelo subyacente a lo que entendemos por computación y que es aplicable a los tres niveles: conocimiento, simbólico y físico. Teniendo en cuenta que en cada nivel la causalidad es diferente, dicho modelo subyacente ha de ser abstracto y aplicable a cada nivel. La obtención de dicho modelo para un problema determinado permitiría obtener el programa correspondiente a un modelo de conocimiento dado.

3 Planteamiento del método de resolución

El modelo abstracto en términos del cual puede definirse un modelo subyacente a la computación parte de la descripción característica de un sistema en términos de sus variables de estado, pero añadiendo e incluyendo explícitamente en dicha descripción la de las condiciones que hacen que, en cada caso, el sistema sufra unas transformaciones en vez de las otras posibles, cubriendo esta descripción todas y cada una de dichas alternativas. Podemos pensar en abstracto en una serie de *r* experimentos con un sistema, de modo que en cada una de dichas series el sistema evoluciona desde un estado inicial a uno final, pasando por n_r estados intermedios, a través de las correspondientes transformaciones, pudiéndose describir por completo dichos estados y transformaciones.

Tras dichos *r* experimentos, si reunimos los resultados obtenidos, observaremos que habrá estados y transformaciones comunes a uno o varios de los experimentos, y otros que serán únicos. En una primera aproximación, podemos representar mediante un grafo los estados y transformaciones, mediante nodos y arcos, respectivamente. El resultado es un tipo de grafo finito, conexo, dirigido, que admite ciclos, y donde hay un nodo (llamémosle inicial) del cual todos los demás son descendientes (incluso puede que él mismo también en algún caso). Esta representación no incluye las condiciones para que el sistema sufra unas transformaciones en vez de otras, que hemos dicho que deberíamos incluir. Así pues, en una segunda y definitiva aproximación, agruparemos de dos en dos los nodos de la primera aproximación junto con el arco que les une, como en un mayor detalle de un nodo de esta segunda aproximación (vemos un mayor detalle, como en un zoom óptico) (ver figura 1). Los nodos de esta segunda aproximación se unen mediante arcos que representan las condiciones a las que nos estamos refiriendo, que describen por qué el sistema sufre unas transformaciones y no otras, según el caso. Este grafo, aunque conceptualmente es diferente, tiene sin embargo las mismas características que el de la primera aproximación: finito, dirigido, conexo, que admite ciclos, con un nodo del que todos los demás son descendientes.

Un grafo del tipo mencionado puede describirse, a su vez, en términos de unos enunciados muy sencillos, ateniéndonos al siguiente modo de proceder:

1. Dar nombre al grafo.
2. Indicar cuál es el nodo inicial.
3. Para cada nodo, indicar cuáles son sus sucesores, en el orden deseado.
4. Para cada sucesor, indicar por qué arco se alcanza.
5. Definir sólo un nodo inicial (aunque haya varios candidatos).
6. Es obligatorio que todos los nodos, salvo el inicial, sean sucesores de algún otro (el inicial puede serlo o no), aunque puede haber nodos que sean sucesores de varios.

De este modo, obtendríamos, por ejemplo, algo como:

Desde el nodo “S^k_1”, se alcanza “S^k_2” por “A^k_2”. Desde el nodo “S^k_2”, se alcanza “S^k_3” por “A^k_3”, se alcanza “S^k_4” por “A^k_4”, se alcanza “S^k_5” por “A^k_5”.

Desde el nodo “S^k_5”, se alcanza “S^k_2” por “A^k_1”.

Desde el nodo “S^k_3”, se alcanza “S^k_6” por “A^k_6”.

Desde el nodo “S^k_4”, se alcanza “S^k_7” por “A^k_7”.

donde S^k_i representa el nodo *i* del grafo *k*, mientras que A^k_j representa el arco *j* del mismo grafo. Esta descripción, como resulta bastante obvio intuitivamente, se obtiene recorriendo el grafo "primero en amplitud", recorriendo los descendientes de cada nodo una sola vez. Nótese que el grafo correspondiente a este ejemplo tiene un ciclo.

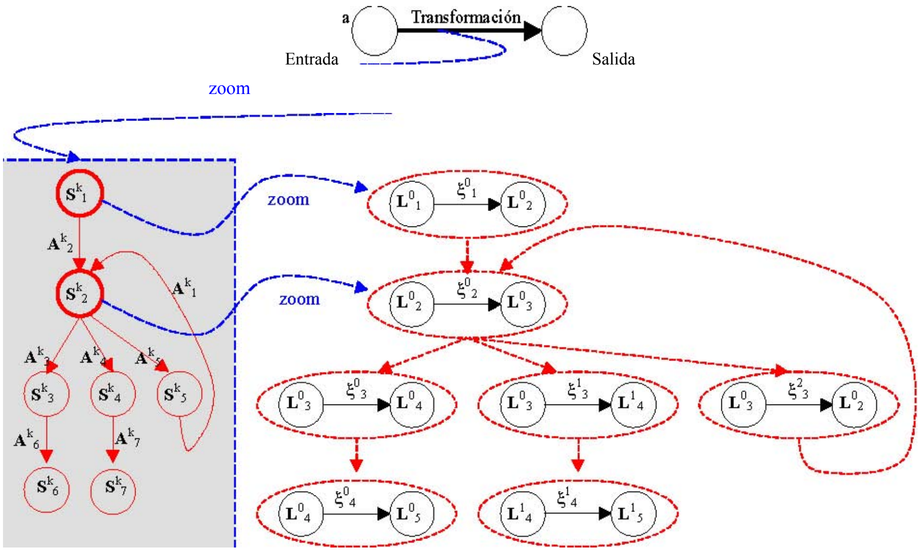


Figura 1. Cada operación de zoom muestra una representación en mayor detalle con información que no es considerada en una representación a un detalle menor. Arriba: grafo representando la computación como un sistema. Izquierda (recuadro azul): detalle de la transformación (segunda aproximación). Derecha: elementos del grafo de la primera aproximación, tal como entran a formar parte de la segunda aproximación. $\mathbf{L}^r_{j+1}$ es la lista de variables de estado con sus valores tras cada transformación ξ^r_j (r = 0, 1, 2). La propia computación es básicamente representada mediante un grafo con dos nodos, que representan la entrada al sistema y la salida del sistema, y un arco representando la transformación total del sistema que, dada una entrada, produce una salida.

La relación con los niveles computacionales se hace a través de la puesta en correspondencia de estos elementos abstractos con los propios de cada nivel. Tomemos en primer lugar el caso simbólico, para un lenguaje de tercera generación. Equiparemos las variables de estado con las del programa, por lo tanto, equiparamos el programa con el sistema; entonces las condiciones que hacen que el "estado" del programa varíe de un modo en vez de otro, resultan venir expresadas en las sentencias "if". Algo similar puede hacerse para otros lenguajes, desde Fortran a Java, pasando por C++, e incluyendo también Prolog y Lisp.

Para el caso del nivel de conocimiento, la correspondencia se obtiene abstrayendo la tarea a resolver computacionalmente. Si tomamos el caso habitual del diagnóstico, por ejemplo como lo describió Chandrasekaran [12,14], consideraríamos que el estado del sistema viene dado por cada uno de los refinamientos sucesivos del diagnóstico; y las condiciones que hacen que dicho sistema varie de un modo en vez de otro

vienen dadas por las características que se tienen en cuenta para derivar por un refinamiento en vez de por otro. Esto es igualmente aplicable, por ejemplo, a una clasificación de tipo jerárquico, a un diagnóstico sistemático, etc. Pero además, en este caso del nivel de conocimiento podemos establecer una correspondencia con los propios enunciados que describen el grafo, equiparando sus elementos con los del lenguaje natural empleado para describir el modelo de conocimiento (dichos elementos del lenguaje realmente resultan en una degradación de la semántica del lenguaje natural y del metalenguaje de la medicina en el sentido en el que se emplea este término en la Física):

> Establezco infección como: el paciente tiene fiebre y ...
> Refino infección en: vírica, bacteriana.
> etc...

Así pues, dados los enunciados que representan el modelo de conocimiento, alcanzamos el nivel simbólico a través del modelo subyacente, estableciendo las pertinentes correspondencias.

4 Una misma estructura subyacente para modelar tareas y ontologías

Es bien conocido que las condiciones en cualquier modelo suelen ser descritas en términos de la lógica, para la cual existen representaciones bien establecidas, tales como la de Lukasiewicz (o "polaca"), entre otras. Todas ellas admiten una representación mediante árboles, aparte de la propia algebraica. Cabría preguntarse ¿puede la computación ser representada mediante una expresión y puede dicha expresión a su vez ser representada mediante un árbol (en vez del tipo de grafo en el que hemos visto que puede inmediatamente ser representada)? La respuesta es sí.

Si consideramos la representación subyacente del apartado anterior, podemos plantear de inmediato una expresión como la siguiente:

$(S^k_1, S^k_2, A^k_2, S^k_2, S^k_3, A^k_3, S^k_4, A^k_4, S^k_5, A^k_5, S^k_5, S^k_2, A^k_1, S^k_3, S^k_6, A^k_6, S^k_4, S^k_7, A^k_7)$

Hoy en día, gracias a la existencia de XML, podemos comprender más fácilmente cómo esta expresión puede representarse mediante un árbol (los enunciados que describen el grafo nos ayudan a encontrar los tags de XML). Para abreviar, esbozamos el inicio del documento XML correspondiente:

<grafo>
<desde>
<nodo>"S^k_1"</nodo>
<se_alcanza> <nodo>"S^k_2"</nodo> <por>"A^k_2"</por>
</se_alcanza>
</desde>
<desde>
<nodo>"S^k_2"</nodo>
<se_alcanza>

<nodo>“S^k_3”</nodo>
<por>“A^k_3”</por>
</se_alcanza>
<se_alcanza>
<nodo>“S^k_4”</nodo>
<por>“A^k_4”</por>
</se_alcanza>
...

Y es sabido que un modelo de documento XML es representable gráficamente mediante un árbol. Si los enunciados del modelo abstracto subyacente, son representables mediante un árbol, por una parte, lo es el propio modelo subyacente, existiendo un árbol equivalente al tipo de grafo mencionado anteriormente (que no era un árbol en general); y por otra parte, el propio modelo a nivel de conocimiento, en base a los enunciados correspondientes a dicho nivel, también es representable mediante un árbol. Esto nos acerca a la representación que mediante árboles hizo el propio Chomsky [14] del lenguaje natural. Pero los grafos árboles no solamente pueden modelar las tareas, como hemos visto hasta ahora. También sirven como modelo subyacente de ontologías ya sea a través de XML (software comercial como TIBCO Designer, o de investigación, como OntoSchema) u otros paradigmas (marcos, o clases-instancias si se prefiere, como en Protégé).

5 Ejemplo de aplicación al diagnóstico en oftalmología

Para terminar con la exposición, se va a mostrar un ejemplo aplicado a un caso de diagnóstico, en concreto el de los problemas genéricos en los ojos, en base a los datos del proyecto DIAGEN. La tarea genérica modelada es la de diagnóstico por clasificación jerárquica, como la definió Chandrasekaran [15], puede considerarse equivalente a la tarea genérica de refinamiento sistemático de KADS [2].

Los enunciados que describen el modelo de tarea genérica (reutilizable), se esquematizan, por lo tanto, en los verbos fundamentales “Establezco” y “Refino”, en base a los cuales se construye el modelo como puede verse en las figuras 2 y 3. La aplicación software (OntoSchema) edita los enunciados conforme a un esquema definido en función de los verbos pertinentes, y en base a dichos enunciados se construye el modelo de conocimiento. Pero al mismo tiempo, la aplicación elabora o “reconoce” el modelo subyacente, de modo que el resultado final es el código fuente en alguno de los lenguajes y entornos más usados y otros tradicionales, tales como Prolog, Lisp, “visual” C++ y Java, Smart Elements (sistema basado en reglas), Fortran... cada uno de cuyos códigos resultantes es directamente compilable y/o ejecutable respectivamente por dichos entornos (el programa ejecutable dialoga con el usuario). La aplicación también puede ella misma realizar una ejecución “on-line”, a modo de intérprete, llevando a cabo el mismo tipo de diálogo que en los casos anteriores.

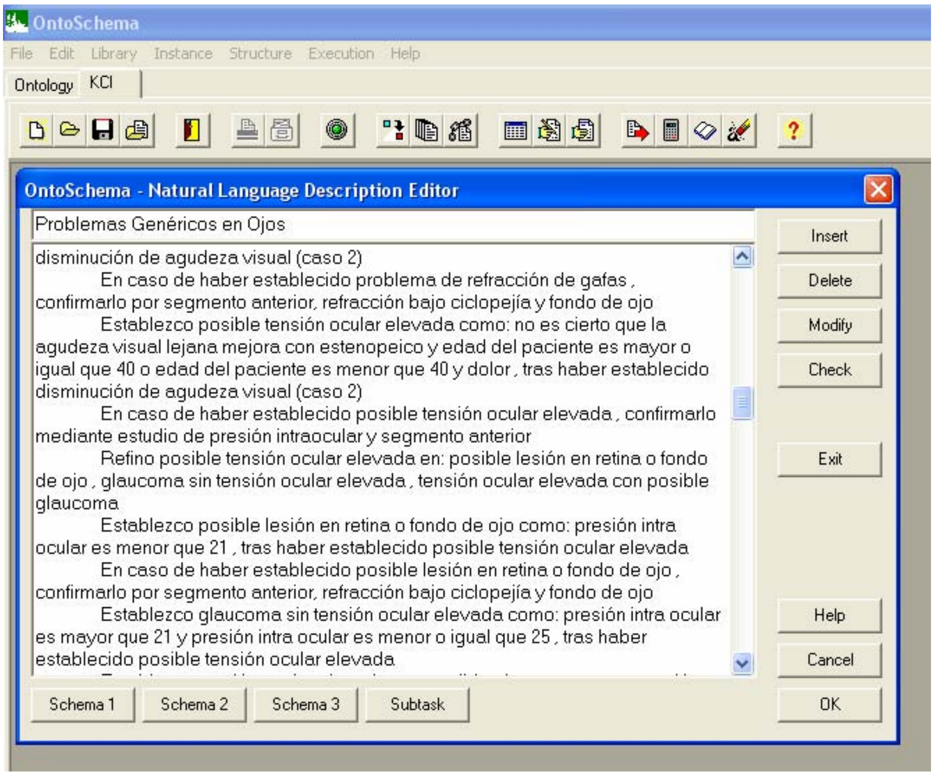


Figura 2. OntoSchema: descripción del modelo de conocimiento de tarea genérica para el diagnóstico de problemas genéricos en los ojos.

Un breve fragmento de la descripción en base a los enunciados concretos de este modelo es como sigue:

...

Establezco disminución de agudeza visual (caso 1) como: la agudeza visual lejana es buena y la agudeza visual próxima es mala, tras haber establecido indicio de disminución de agudeza visual

Refino disminución de agudeza visual (caso 1) en: presbicia, hipermetropía

Establezco presbicia como: edad del paciente es mayor o igual que 40, tras haber establecido disminución de agudeza visual (caso 1)

En caso de haber establecido presbicia, confirmarlo mediante estudio refractivo con lentes positivas

Establezco hipermetropía como: edad del paciente es menor que 40, tras haber establecido disminución de agudeza visual (caso 1)

En caso de haber establecido hipermetropía, confirmarlo por segmento anterior, refracción bajo ciclopejía y fondo de ojo

...

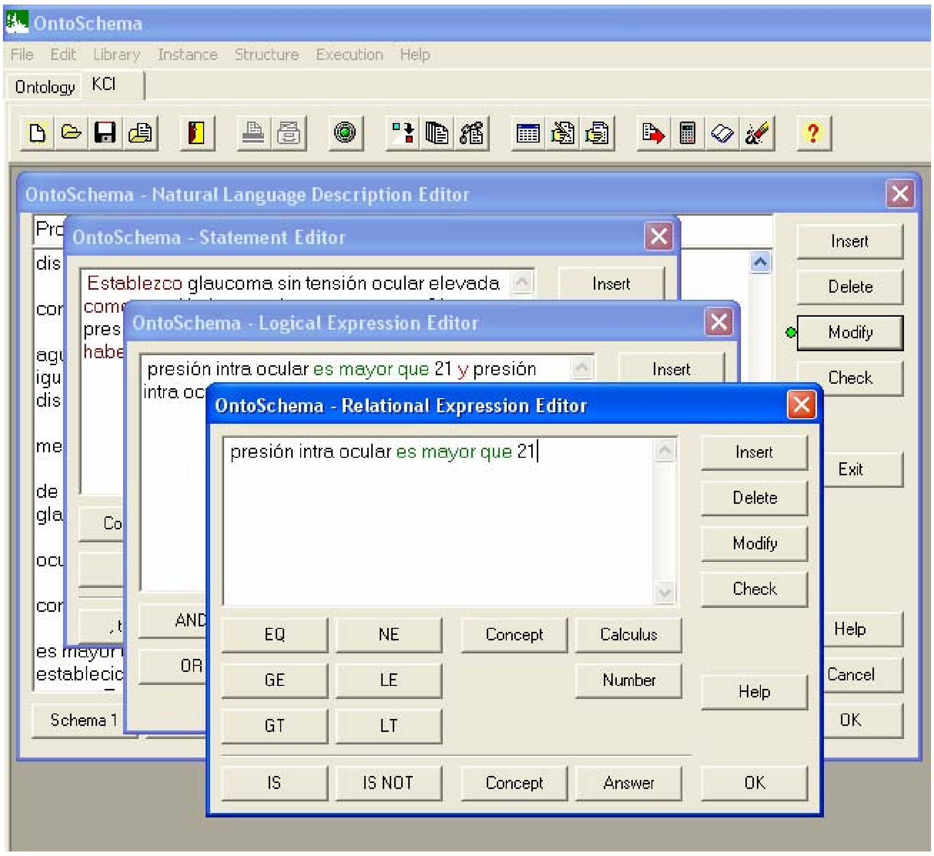


Figura 3. OntoSchema: Despliegue de las ventanas de edición de los enunciados correspondientes a la figura 2 y algunos de sus diferentes componentes. El enunciado elegido en esta imagen en particular es "Establezco glaucoma sin tensión ocular elevada como: presión intra ocular es mayor que 21 y presión intra ocular es menor o igual que 25, tras haber establecido posible tensión ocular elevada"

La ontología correspondiente a este ejemplo puede verse en la figura 4, editada también mediante OntoSchema.

Este desarrollo de un sistema de diagnóstico para oftalmología usando nuestro entorno OntoSchema tiene el carácter de prototipo al que hay que seguir completando al menos en los siguientes puntos: (1) el método sistemático y jerárquico propuesto ("es-tablece-refina") debe completarse para que el sistema pueda contemplar un espectro más amplio de situaciones patológicas (además de agudeza visual, ojo rojo y ojo seco) y de relaciones ente las distintas patologías de forma que sea posible analizarlas de forma conjunta y diferencial. (2) Incorporando nuevos métodos (nuevos verbos inferenciales además de establecer y refinar y nuevos esquemas de conectividad entre verbos) y nuevas tareas que nos permitan modelar de forma más completa el conjunto de actividades usuales en un servicio de oftalmología.

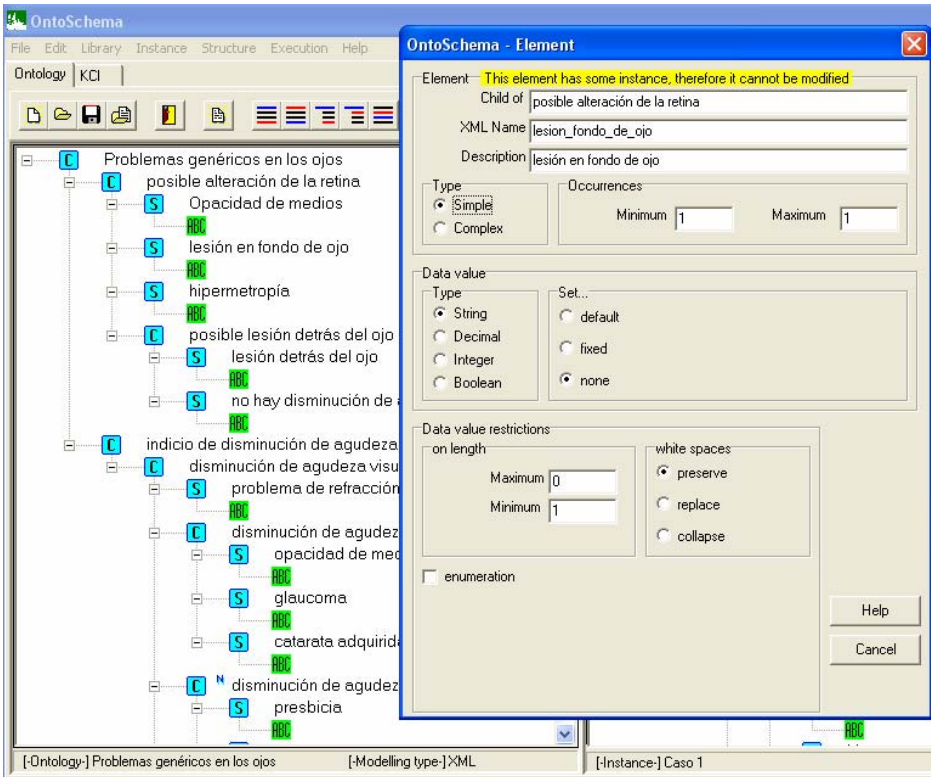


Figura 4. OntoSchema: ontología para un modelo de problemas genéricos en los ojos, ventana del fondo a la izquierda. La estructura subyacente es un árbol, en particular modelado mediante XML, como puede observarse en los items de la ventana de edición superpuesta y también se indica al pie de la ventana del fondo. Al fondo a la derecha, no es visible, una de las instancias del modelo.

6 Conclusiones

Hay una preocupación general en nuestro grupo de investigación en explorar de forma sistemática un conjunto de procedimientos concurrentes o alternativos que nos permitan construir sistemas basados en conocimiento (SBC) para la tarea del diagnóstico en medicina a partir de las descripciones en lenguaje natural que un médico especialista hace acerca de cómo cree él que resuelve esa tarea de diagnóstico en su actividad clínica diaria.

En este trabajo hemos descrito un procedimiento efectivo de generar código ejecutable en un lenguaje de programación de alto nivel a partir de una reescritura estructurada en esquemas del modelo a nivel de conocimiento de la tarea de diagnóstico. Hemos ilustrado este procedimiento aplicándolo al dominio de la oftalmología pero creemos que no es difícil extrapolarlo a otras especialidades médicas. Al compartir el

modelo y el programa una misma estructura formal subyacente (un grafo jerárquico recursivo y recurrente) se facilita el seguimiento de los cambios de semántica que se han efectuado en la formalización y programación del modelo. Además se facilita también el uso medido y preciso de estas mismas tablas de semántica al interpretar los resultados del cálculo para cada sesión clínica de cada enfermo concreto.

Adicionalmente, estamos diciéndoles a los médicos que si en algún momento quieren aproximar el arte del diagnóstico a la ciencia e ingeniería convencionales, puede serles de gran ayuda el descubrir y hacer explícito el modelo formal subyacente a su razonamiento.

Finalmente, nuestro objetivo a largo plazo es aumentar el repertorio de tareas y métodos para los que disponemos de un modelo formal subyacente e integrarlos en una metodología de desarrollo de SBCs en diagnóstico médico. La riqueza, variabilidad, expresividad y sutileza semántica del lenguaje natural nos hace intuir que nuestro objetivo es de largo alcance. Sin embargo creemos que hemos iniciado el camino.

Agradecimientos

Uno de los autores (J. Mira) agradece la subvención económica recibida del proyecto TIN2004-07661-C02-01 durante la realización de este trabajo.

Referencias

1. Herrero, J.C. Un modelo de correspondencias entre el nivel de conocimiento y el nivel simbólico para un conjunto de tareas genéricas. Ph D Tesis (UNED, 1998)
2. Tansley, D.S.W., Hayball, C.C. Knowledge-Based Systems Analysis and Design. A KADS Developper's Handbook. (Prentice-Hall, 1993)
3. Wielinga, B.J. y A.Th. Schreiber, A. Th. Knowledge Technology: Moving into the next Millenium. En Methodology and Tools in Knowledge-Based Systems, IEA/AIE-98 Vol I, 1-20. Mira, del Pobil y Ali (Eds.) (Springer, 1998)
4. Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., Van de Velde, W. and Wielinga, W. Knowledge Engineering and Management, The CommonKADS Methodology. (The MIT Press, 1999)
5. http://protege.stanford.edu/
6. Herrero, J.C. and Mira, J. In Search of a Common Structure Underlying a Representative Set of Generic Tasks and Methods: The Hierarchical Classification and Therapy Planning Cases Study. In Mira, del Pobil & Ali (eds.) Methodology and Tools in Knowledge Based Systems, pp. 21-36. (Springer, 1998)
7. Herrero, J.C. and Mira, J. SCHEMA: A Knowledge Edition Interface for Obtaining program Code from Structured descriptions of PSM's. Two cases study. Applied Intelligence 10 (2/3), (1999) pp. 139-153.
8. Taboada, M., Des, J., Mira, J. and Marin, R. Development of diagnosis systems in medicine with reusable knowledge components. IEEE Intelligent Systems, 16 (2001), 68-73.
9. http://www.tibco.com/
10. Newell, A. The Knowledge Level: Presidential Address. AI Magazine 2 (2) (Summer 1981) 1-20, 33.

11. Mira, J., Herrero, J.C. and Delgado, A.E. Where is Knowledge in Computational Intelligence? On the Reduction of the Knowledge Level to the Level Below. Proceedings of the 24th Euromicro Conference, pp. 723-732. (IEEE, 1998)
12. Chandrasekaran, B. Generic Tasks in Knowledge-Based reasoning: High-level Building Blocks for Expert System Design. IEEE Expert 1 (Fall 1986) 23-30.
13. Herrero, J.C. XML y el manejo en Internet de conocimiento estructurado. En Mira (Ed.) Conocimiento, Método y Tecnologías en la Educación a Distancia, pp 110-116. (UNED, 2000)
14. Chomsky, N. Language and Mind. (Harcourt Brace Jovanovich Inc., 1968)
15. Chandrasekaran, B., Johnson T.R, and Smith, J.W. Task Structure Analysis for Knowledge Modeling. Communications of the ACM 35 (9) (September 1992) 124-136.
16. Herrero, J.C. and Mira, J. Causality Levels in SCHEMA: A Knowledge Edition Interface. (IEE Proceedings-Software, 2000). Vol 147, No 6, pp 193-200

Aprendiendo a aprender: De máquinas listas a máquinas inteligentes

Bogdan Raducanu[1] y Jordi Vitrià[1,2]

[1] Centre de Visió per Computador, Edifici O - Campus UAB
[2] Departament de Cienciès de la Computació, Universitat Autònoma de Barcelona
08193 Bellaterra, Barcelona, España
{bogdan, jordi}@cvc.uab.es

Resumen Desde su aparición, hace más de cinco decadas, uno de los retos más ambiciosos de la Inteligencia Artificial ha sido la creación de sistemas computacionales inteligentes. A pesar de los impresionantes avances alcanzados, todavía estamos lejos de tener máquinas que se acerquen al nivel de la inteligencia humana. Esto se debe al hecho de que los sistemas de hoy en dia carecen de sentido común, un elemento característico de las personas. El presente artículo está enfocado sobre el análsis de las distintas estrategías de aprendizaje, tipos de contexto que intervienen en el proceso de aprendizaje y su aplicación en la interacción persona-máquina.

Palabras clave: sistemas inteligentes, aprendizaje cognitivo, contexto, robótica social, reconocimiento de caras

1. Introducción

Conocimiento no es equivalente a inteligencia. Hoy en dia (debido a los avances en Inteligencia Artficial y potencia de cálculo), las máquinas son capaces de hacer cosas remarcables: hay algoritmos de jugar ajedrez a nivel de grandes maestros, aplicaciones complejas para coordinar el desplazamientos de los efectivos militares en los campos de batalla, herramientas de diseño que nos ayudan desde el desarollo de los circuitos electronicos más sofisticados hasta las aeronaves más complejas. Nadie pone en duda todas estas evidencias. Pero en cambio, a pesar de la complejidad de los sistemas mencionados, ninguno de ellos es capaz de, por ejemplo, interpretar una fotografia, comentar un texto, contestar a una pregunta, cosas que son obvias para cualquier persona. A simple vista, no podemos negar que no existen las herramientas necesarias o que no hayan programadores capaces de desarollar sistemas que presenten un cierto nivel de razonamiento (hay un número sin fín de ejemplos en el ámbito de los sistemas expertos). De hecho, el campo de Inteligencia Artificial está lleno de teorías y algoritmos de aprendizaje. La pregunta que nos planteamos entonces es: que componente, a parte del conocimiento especializado, se ha omitido en la fase de programación,

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 34-45, 2006.

para que una máquina sea realmente inteligente? La respuesta a la pregunta anterior viene dada por el concepto de 'sentido común'. Pero que es el sentido comuún'? La siguiente definición viene dada por Marvin Minsky en [11]: "the mental skills that most people share". Para nosotros, el sentido comun es una característica tan natural, que muchas veces, en la vida cotidiana, lo ignoramos. En cambio, una máquina no tiene ni idea de como somos, que sentimos o cuales son nuestras preferencias.

Un sistema experto, por ejemplo, tiene una representación del conocimiento específico (premisas y reglas de inferencia) para solucionar un problema particular. Pero para dotarlo de sentido común, hace falta alimentarle con información mucho más general: conocimientos de física para saber como se comportan los objetos, conocimiento social para entender como interaccionan las personas, conocimientos de psicología, para entender como funciona la mente humana, etc. Para cada categoría de conocimiento, las personas utilizamos distintos modos de representación del mismo y distintas estrategías de razonamiento. Por lo tanto, dotar una máquina de sentido común no es especificarle que razonamiento aplicar sobre un cierto conjunto de datos (esto ya está solucionado), sino como seleccionar la estrategía de razonamiento adecuada y como seleccionar sobre que subconjunto de conocimiento (de la multitud de datos representados) debe aplicarla. Las complicaciones que aparecen aquí son aún mas grandes. En primer lugar, dado un conjunto tan grande de conocimiento general (hechos), ¿cuál es el modelo de representación adecuado? ¿Es posible que el cerebro humano utilice varias representaciones para el mismo hecho? En segundo lugar, tenemos que disponer de métodos para definir y representar el conocimiento funcional: las distintas estrategias de razonamiento, planificación, predicción, etc. Una discusión más detallada sobre estos aspectos está fuera del proposito de este trabajo, pero se puede encontrar en [17].

El artículo está estructurado de la siguiente manera: en la sección 2, hacemos un repaso de las teorías de aprendizaje existentes. La sección 3 está dedicada a la presentación del papel del contexto en el proceso de aprendizaje. En la sección 4 hablaremos de un estudio que está en marcha, de la aplicación del aprendizaje cognitivo en el entorno de la robótica social (como exponente del área de interacción persona-máquina). Finalmente, sección 5 contiene nuestras conclusiones y direcciones de trabajo futuro.

2. Teorías de aprendizaje

El término de aprendizaje, dentro de la Inteligencia Artificial, se refiere a la habilidad de una máquina de adquirir ciertos conocimientos de tal manera que, debido a los cambios en la representación interna de los datos (como consecuencia de nuevas experiencias), mejora su funcionamiento/comportamiento con el tiempo. Las preguntas que nos podriamos plantear son las siguientes: ¿Porque queremos que las máquinas aprendan? ¿Porque no diseñamos desde un principio, una máquina capaz de tener la funcionalidad deseada? Las respuestas a estas preguntas vienen desde varias direcciones: algunos conocimientos solo se

pueden adquirir basandonos en ejemplos, puede haber una relación estricta entre los datos de entrada y salida (o que los datos presenten una correlación oculta), el entorno de trabajo de la máquina cambia con el tiempo o incluso nuevos conocimientos son adquiridos por parte de los programadores. El aprendizaje converge de varias áreas: estadistica, modelos cognitivos, teoría de control adaptativo, modelos psicologicos, etc. Más detalles sobre estos temas se pueden encontrar en [13].

En conclusión, y tal como vamos a ver con más detalles a continuación, hay una relación muy estrecha entre la adquisición de los datos, representación del conocimiento y estrategia de aprendizaje.

2.1. Aprendizaje basado en representación simbólica

En la metodología tradicional de la Inteligencia Artificial, la resolución de problemas se basaba en la decomposición funcional [8] y abstracción de los datos mediante una representación simbólica [12]. Los programadores partían de un conjunto de premisas e intentaban construir un modelo del mundo. Cuando se solicitaba una respuesta por parte del sistema, se intentaba encontrar una correspondencia uno-a-uno entre los datos de entrada y el motor de inferencia, que constituía la base de conocimientos del sistema. Por lo tanto el razonamiento se basaba en el ciclo 'percibir-interpretar-responder'.

El error que se cometía con la abstractización de los datos era que se utilizaban unos simbolos que no tenían nada que ver con el mundo real. Estos símbolos existían unicamente en la visión del programador. Con la abstracción de los datos, para el programador era imposible prever todos los casos posibles que podrían aparecer durante el funcionamiento del sistema. La explicación era que de este modo, se podría reducir sustancialmente los problemas a resolver. La arquitectura para este tipo de sistemas, era centralizada (de aquí la sintagma "brain-in-a box") y por lo tanto ocurría muy a menudo que el sistema quedaba bloqueado, cuando recibía un estimúlo cuya respuesta no estaba prevista en su diseño. Un ejemplo clásico es representado por el MYCIN [16], un sistema experto para el diagnóstico de las infecciones bacterianas. El sistema no tenía ningún modelo sobre lo que es una persona o que le puede pasar. Si le decías por ejemplo que el paciente se ha hecho un corte y pierde sangre, el sistema intentaba determinar la causa bacteriana que pudo causar este problema.

La explicación para el fracaso de los sistemas basados en representación simbólica podría ser que el cerebro humano representa la información no solamente por su categoría, sino también por la modalidad con la cual ha sido obtenida. En otras palabras, la representación está basada también en los mecanismos sensoriales que han contribuido a la adquisición de la información [5].

2.2. Aprendizaje basado en comportamientos

Visto el aparente fracaso en el desarollo de sistemas inteligentes mediante la abstracción del conocimiento, era obvio que un nuevo paradigma era necesario. En [4], el autor propuso el desarollo de sistemas inteligentes que no necesitaban

una representación centralizada del mundo. La idea era de que el aprendizaje pueda tener lugar por el emparejamiento directo entre los datos de entrada (estimulos) y sus respuestas (el llamado "sensation-to-action"). En este caso, la arquitectura general del sistema era distribuida, representada por varios modulos que desarollaban un comportamiento muy sencillo. Los comportamientos de alto nivel se realizaban como consecuencia de la combinación de estos comportamientos simples. Esta idea también ha sido utilizada por Minsky en [11] para explicar la emergencia de la inteligencia humana. En su visión, la mente está formada por un conjunto de 'agentes' que compiten y cooperan entre ellos.

Al ser excluída la representación explicita de los datos, no se puede hablar de un método para medir directamente la capacidad de conocimiento del sistema. En cambio, el nivel de aprendizaje se evalúa por el análisis de su comportamiento. En otras palabras, se puede decir que el sistema aprende cuando muestra un cambio en su comportamiento. Como consecuencia de ello, un requisito fundamental para estos sistemas, es representado por su instalación directa en el entorno donde iban a funcionar. Esto es debido al hecho de que su comportamiento se desarrolla basandose en la interacción directa con el entorno. Detrás de ello se encuentra la idea filosófica sobre el dualismo cuerpo-mente. Este es un concepto muy conocido en el campo de la Inteligencia Artificial y se refiere al hecho de que la mente no puede imaginar mas alla de lo que está percibido por el sistema sensorial. Esta realidad está metaforicamente descrita en [10] por el sintagma: "the body is the anchor of the mind".

2.3. Aprendizaje cognitivo

El modelo de aprendizaje presentado en la sección anterior tiene una desventaja: está limitado en gran parte solo a los cambios del entorno, y en menor medida a la interacción entre los componentes internos. Por lo tanto, a veces es dificil de establecer si un cambio interno se ha debido realmente a la modificación de las condiciones externas. Por otro lado, hay sistemas cuyas características se parecen muchisimo a los descritos anteriormente, pero de los cuales no se puede decir que son inteligentes. Como contraejemplo, se puede mencionar un sistema cibernético. Consideremos un sistema de aire acondicionado. Es verdad que debido a la arquitectura interna (el termostato) este sistema cambia su comportamiento como consecuencia de la modificación de las condiciones externas (temperatura). Pero está muy lejos de ser considerado un sistema inteligente.

Por lo tanto, los investigadores han llegado a la conclusión de que el aprendizaje no se puede relacionar únicamente con los cambios comportamentales, sino que es también necesario estudiar los cambios internos que tienen lugar en la representación del conocimiento [14]. Con este nuevo paradigma, el aprendizaje puede tener lugar sin que se note un cambio aparente en el sistema. El aprendizaje cognitivo consiste en un proceso integrado, recursivo, que tiene como fín construir un modelo del mundo y una adaptación contínua de este modelo. Por lo tanto, es el mismo sistema el responsable de como analizar, interpretar y modelar el mundo. El sistema aprenderá nuevos conceptos (desarrollará nuevas competencias) basándose en el conocimiento aprendido y en

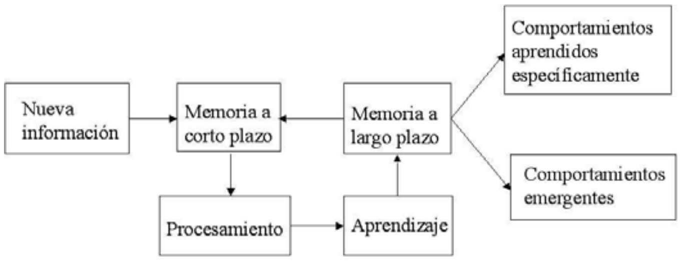


Figura 1. Modelo de aprendizaje cognitivo. Para más detalles, ver texto.

la experiencia adquirida. Con la llegada de nueva información, esta deberá ser analizada y el sistema tomará la decisión de, si es información útil, integrarla en la representación existente. A veces, se puede incluso llegar a cambiar la estructura de la representación como consecuencia de esta nueva información. Como respuesta, el sistema podrá desarollar dos clases de comportamiento: una clase de comportamientos aprendidos especificamente y otra clase de comportamientos emergentes. Este modelo de aprendizaje cognitivo está representado en la figura 1.

Una característica muy importante cuando hablamos del aprendizaje cognitivo es el acceso a la información acumulada. Hay que destacar dos tipos de memorias: una a corto plazo y otra a largo plazo. La memoria a corto plazo se refiere a la información que se necesita para finalizar una tarea especifica (con un propósito y una duración muy bien definidas). Después, la información se puede eliminar (sin que afecte de algun modo la viabilidad) para no ocupar innecesariamente los recursos del sistema. Por otro lado, la memoria a largo plazo es la información necesaria que le permite al sistema existir y funcionar en el tiempo (medido en años por ejemplo). Un problema que puede ocurrir con la memoria a largo plazo, es el proceso de 'olvido' o de 'degradación' (perdida parcial de las características asociadas con una cierta información). El fenómeno de olvido es una propriedad fundamental para el funcionamiento del cerebro humano.

Un elemento muy esencial en el proceso de aprendizaje cognitivo está representado por el contexto: el conjunto de factores que son determinantes en la adquisición, representación y (más tarde) recuperación de la información. En el siguiente párrafo vamos a introducir la noción de contexto y como se representa en distíntos áreas.

3. Contexto

Según [23], el término de 'contexto' tiene sus fundamentos en lingüistica. La palabra viene compuesta por las particulas 'con' y 'texto' y se refiere al significado extraído de la lectura de un texto. Hoy en día, la acepción del término es mucho más amplia, y se refiere a un marco particular en el proceso de comunicación, basado en unas experiencias comúnes.

Con el paso del tiempo, la utilización del término 'contexto' se ha especializado en distintas áreas de conocimiento. En [3] se puede encontrar una discusión muy detallada sobre este tema. A continuación, vamos a destacar la utilización del 'contexto' solo en las áreas de interés.

En Inteligencia Artificial, viene asociado con el 'frame problem': representar en el lenguaje de la lógica el resultado de una acción de forma implícita, sin recurrir a una representación explícita de los efectos que no es han producido como consecuencia de la acción. Es uno de los problemas más difíciles con el cual se ha confrontado la Inteligencia Artificial.

En el ámbito de la comunicación (escrita o verbal), el 'contexto' se refiere a unas propiedades del proceso de interacción entre varios agentes. Esta interpretación es opuesta a la del 'contexto' como propriedades de un fenómeno en particular. En otras palabras, la noción de 'contexto' no existe sin interacción. El 'contexto' es considerado como una 'história' de todo lo que pasó en un cierto periodo de tiempo (desde que se inició la interacción), el conjunto de conocimientos de que disponen los agentes y de las particularidades derivadas del tema sobre el cual se están enfocando en un cierto instante. De este modo, el contexto aparece como un 'espacio de conocimiento compartido'. Como consecuencia, se pueden desarollar heramientas (asistentes) cuyo fin es predecir y anticipar las solicitudes de los agentes. En el caso de la comunicación escrita, los ejemplos vienen dados por los asistentes asociados a los editores de texto: nada más introducir las primeras letras de una palabra, el asistente busca en su diccionario las posibles opciones para poder continuar. En el caso de la comunicación verbal, teniendo información sobre el tema de la conversación podemos eliminar la incertidumbre provocada por una expresión ambigua, implícita de los hechos. Por ejemplo, en la siguiente frase: 'He visto un gato por la calle', al no conocer el contexto, no se puede inferir el significado. No se sabe si la persona se refiere al gato como animal, o al gato como herramienta para coches.

En el ámbito de la visión, el termino 'contexto' se refiere tanto al entorno que rodea un objeto en particular, como a las propiedades intrínsecas del mismo objeto. Incluso hoy en día, a pesar de los avances realizados en el área de procesamiento de imágenes y el reconocieminto de patrones, la identificación visual de objetos queda un problema parcialmente solucionado. Visto desde el punto de vista de la Inteligencia Artificial, la tarea de describir una escena (en término de los objetos que la componen) es uno de los hitos más ambiciosos que han quedado por resolver. Entre los factores que dificultan el reconocimiento visual se pueden mencionar las condiciones de iluminación, los cambios en la apariencia (vistas desde diferentes angulos, deformaciones, transformaciones lineales), posibles oclusiones, etc. Por esta razón, el uso del contexto pretende simplificar el proceso de reconocimiento por parte del sistema visual. Con el uso del contexto, no solamente se puede conseguir la eliminación de muchos errores, sino también la ambigüedad que puede dificultar a veces la toma de decisiones.

Experimentos en análisis de escenas [2] han confirmado que el sistema visual humano utiliza de manera extensiva el contexto para facilitar la detección y el reconocimiento de objetos. Además, en el mundo real existe una relación muy

estrecha entre un objeto y el entorno en el cual está situado. Por lo tanto, la decisión sobre la presencia o la ausencia de un objeto en la escena está en gran medida influenciada por ella; la presencia de distintos tipos de objetos puede estar fuertemente correlacionada: por ejemplo, si se consigue detectar un monitor en una imagen, se puede esperar a que se encuentre también un teclado. Parece que el sistema visual lo primero que hace es un análisis global de la escena para poder estimar los objetos que pueden aparecer en ella. El contexto puede ayudar a la identificación de los objetos presentes en la escena de dos modos [18],[19] : en primer lugar, cuando las características de los objetos son parcialmente observables o están afectadas por ruido; en segundo lugar, asumiendo que la identificación se ha producido satisfactoriamente, el contexto todavía puede ayudar a la hora de eliminar las posibles incertidumbres creadas en la fase de clasificación del objeto. Además, los mismos autores proponen una representación del contexto teniendo en cuenta las características generales que aparecen en la imagen (realizar una representación de la imagen con un número reducido de dimensiones, por ejemplo). Esta tranformación se puede realizar de una manera relativamente sencilla, sin la necesidad de identificar unas regiones en concreto en la imagen. Con esta representación luego, se facilita la detección de objetos individuales, porque el contexto puede dar indicios muy valiosos sobre la posición y el tamaño de los objetos en la imagen.

4. AiboFace: un estudio para el desarollo del aprendizaje cognitivo en robots

Desde su creación, los sistemas computacionales han sido siempre enfocados hacia la máquina y no hacia las personas. Hasta hoy en dia, la interacción persona-máquina suponía que el usuario debe comprender el funcionamiento de las máquinas, como 'piensan' y como están construidas. Como consecuencia, estamos obligados a trabajar en sus términos, utilizando su lenguaje y unos dispositivos especificos para comunicarnos con ellas (ratón, teclado, etc.). Con la aparición de la realidad virtual, la cosa era aún más absurda: estamos obligados a sumergirnos en un mundo sintético, creado por ellas.

Pero en el futuro, los investigadores auguran unos cambios fundamentales en la interacción persona-máquina. En el futuro, la interacción será centrada en el usuario. En otras palabras, el usuario no tendrá porque estar preocupado por la presencia de las máquinas o como funcionan. En cambio, las máquinas estarán previstas con capacidad sensorial, que les permita identificar la presencia de una persona y estar atenta en cada momento a sus acciones, pero al mismo tiempo respetando nuestra privacidad e intimidad. Como consecuencia, ellas estarán omnipresentes en nuestras vidas, rodeandonos en el trabajo y en el hogar. De hecho, esta realidad ha sido anticipada por Mark Weiser en [22].

Dentro de esta visión, un lugar destacado está representado por la robótica social. En la robótica social, los investigadores consideran los róbots como unas plataformas para formalizar y probar las capacidades cognitivas de los humanos. Al mismo tiempo, por la implementación de estos modelos, podemos tener una



Figura 2. El robot AIBO de SONY con sus dos juguetes favoritos: la pelota y el hueso.

mejor comprensión de su funcionamiento en las personas. Ultimamente, se intenta extender el uso de estos robots para convertirlos en verdaderos asistentes para personas de la tercera edad. Experimentos psicológicos y sociales han demostrado que pueden ser de gran ayuda en levantar el estado de ánimo de las personas [21], pero también de avisar a los servicios competentes en caso de una situación de emergencia. Por otro lado, se ha estudiado el uso de estos róbots en clínicas pediátricas y las observaciones resultadas de la interacción con los niños han sido analizadas [24]. Debido a la gran complejidad que supone el desarollo de esta clase de aplicaciones, varias áreas estan implicadas: diseño de interfaces, psicología, neurociencias, etc.

Este nuevo paradigma en robótica, contrasta fuertemente con la visión que se tenía sobre los robots hace un par de décadas. Entonces, eran visto solamente como unos sustitutos para las tareas que implicabann acciones repetitivas o que suponían un alto nivel de riesgo para el operador humano. Los primeros robots que se fabricaron eran muy limitados, siendo diseñados para unos entornos muy especificos y siendo capaces de realizar unas tareas muy concretas.

En el caso de los robots sociales, una de las tareas básicas que tienen que solucionar es la detección de personas. Las caras representan de lejos el mejor indicio sobre la presencia de una persona en la vecindad del robot. El argumento para esta afirmación tiene sus raices en la visión biológica. En [7], los autores reclaman el hecho de que los recién nacidos llegan al mundo con una pre-disposición de reconocer caras. Parece ser que, en general están atraídos por patrones en movimiento que tienen estructura parecidas a la cara. Otro argumento presentado en el mismo trabajo, subraya el hecho de que para los humanos nos es más fácil reconocer caras (si están presentadas en posición frontal) que cualquier otro objeto.

Para nuestro estudio, utilizamos un robot AIBO [1] de la marca SONY (figura 2). Creado inicialmente con el proposito de servir como un juguete de entretenimiento más, ha sido rápidamente adoptado por la comunidad científica que ha visto en él uno de los mejores entornos para el desarollo y prueba de las teorías del ámbito de la robótica social (como el aprendizaje cognitivo, por ejemplo). El nombre del robot puede ser interpretado de dos maneras: por un lado, la

palabra 'aibo' significa 'compañero' en japones; por otro lado, su nombre puede ser visto como la combinación entre Inteligencia Artificial (AI - por sus siglas en inglés) y 'roBOt'. AIBO es un robot que viene pre-programado con capacidad para aprender, responder a una variedad de estímulos (reconocimiento de voz, tacto, visión), expresar deseos y mostrar emociones. Sin duda, AIBO es 'algo' distinto. Esto se nota desde el primer instante que se quiere adquirir uno. El personal de venta de SONY te advierte de que un AIBO no se compra, sino se adopta.

Debido a su curiosidad, el AIBO está en un proceso contínuo de aprendizaje y adaptación al entorno [9], [15]. Como consecuencia de ello (a las particularidades con las cuales se han confrontado cada uno de ellos), podemos afirmar que no existen dos AIBOs iguales en el mundo (del mismo modo que no existen dos personas iguales).

A parte de los comportamientos pre-programados, el AIBO viene también con su proprio entorno de desarollo de aplicaciones (unas librerías de funciones para el Visual C++), llamado AIBO Remote Framework. Estas librerías nos permiten desarollar aplicaciones del tipo cliente-servidor, entre el PC y el AIBO, teniendo como 'soporte' una conexión de red inalámbrica entre los dos. El problema que nos planteamos estudiar es el aprendizaje gradual por parte del AIBO de las personas, mediante el reconocimiento facial. Para este propósito necesitamos desarollar unos algoritmos especificos. La aplicación que nos planteamos es relativamente nueva en el ámbito de reconocimiento de caras y se podría expresar como 'reconocimiento no-supervisado'. La idea es la siguiente: al principio, el robot no conoce ninguna persona (el conjunto de aprendizaje es vacío). Con el transcurso del tiempo, a medida que ve nuevas caras, empieza a construir de forma incremental la base de datos.

La estrategía de aprendizaje que se quiere emplear se puede dividir en dos fases, cada una de ellas automatizada. En una primera fase, al robot se le puede 'enseñar' caras, pero no es capaz de diferenciarlas. En una segunda fase, se pretende que el mismo robot desarolle competencias para ser capaz de clasificarlas (del conjunto que el mismo ha adquirido). Durante este proceso de detección/reconocimiento, algunas características de las caras se podrán distinguir claramente de otras, mientrás que algunas se inferirán del contexto. Esta estrategía de aprendizaje es muy similar a la utilizada por humanos. En los primeros meses de nuestra infancia, solo somos capaces de distinguir la clase 'cara' dentro del conjunto general de objetos [6]. Al cabo de un cierto periodo de tiempo, y a medida que nuestras capacidades cognitivas han evolucionado, no solamente somos capaces de identificar las personas, sino también reconocer el género (hombre/mujer) o diferenciarlas según la edad (niño/joven/adulto/anciano).

Al nivel más técnico, el sistema de reconocimiento empleara dos tipos de memoria: una memoria a corto plazo, y una memoria a largo plazo [25]. Con la memoria a corto plazo pretendemos que el robot sea capaz de mantener una coherencia sobre la identidad de la persona a quién esta viendo mientras dure la sesión (una sesión se define como el periodo transcurrido desde la aparición de una persona en la escena, hasta su salida). Con la memoria a largo plazo,

el proposito es de crear una base de datos con los 'amigos' del robot (reidentificación: ¿de que me suena esta cara?). Entre los dos tipos de memoria existe una relación muy estrecha: las nuevas imágenes de cara tomadas durante las sesiones, se añadirán a la memoria a largo plazo, para actualizar de este modo su contenido. Este proceso de 'renovación' tendrá una componente adaptativo en el sentido de que cuando se construya el modelo de cara de una persona, las instancias más recientes tendrán un peso mayor en el cálculo del modelo, que las más viejas. Se llegará a un cierto momento, incluso, que las imágenes más antiguas sean completamente y definiitivamente eliminadas de la base de datos (es la fase de 'olvido' del proceso de aprendizaje).

Además, cada persona tendrá asociada un 'peso' proporcional con el número de veces que ha sido vista. Con ello se pretende que el robot desarolle un comportamiento particular (que muestre 'más interés') hacia las personas que han sido vistas más a menudo (tienen un 'peso' grande), con respeto a personas que han sido vistas menos veces (tienen un 'peso' pequeño).

En estos momentos nos encontramos en la primera fase del proceso de aprendizaje descrito anteriormente: adquisición automática de caras. En concreto, hemos implementado un detector de caras basado en [20]. El detector está relacionado con los circuitos motrices de la cabeza del robot, de tal modo que el efecto conseguido es de un seguimiento activo de la persona: el robot mueve su cabeza en concordancia con el movimiento de la persona en su campo visual, de tal modo que la persona se queda siempre centrada en la imagen capturada por la cámara del robot. En figura 3 mostramos unas instancias del proceso de captura de caras.

5. Conclusiones y trabajo futuro

En este artículo hemos presentado las causas por las cuales los sistemas computacionales de hoy son solamente listos, sin ser inteligentes. El componente que falta en su diseño es el 'sentido común', un rasgo carácteristico de las personas. Despúes de repasar brevemente las teorías de aprendizaje existentes, hemos subrayado la relevancia del contexto en el proceso de aprendizaje. Su uso es imprescindible si deseamos obtener buenas respuestas por parte del sistema incluso en situaciones cuando los datos son incompletos o corruptos. También su papel puede ser fundamental en el proceso de clasificación para la elmininación de la incertidumbre. Finalmente hemos presentado nuestra propuesta de estudio para desarollar el aprendizaje cognitivo en un robot a través de la identificación de personas. En estos momentos, el robot es capaz solamente de detectar caras (y de seguirlas), pero nuestra perspectiva para el futuro es de dotarle con un sistema de reconocimiento no supervisado de caras.

Agradecimientos

El presente trabajo ha sido posible gracias al proyecto MCYT Grant TIC2003-00654 del Ministerio de Ciencia y Tecnología de España. Bogdan Raducanu es

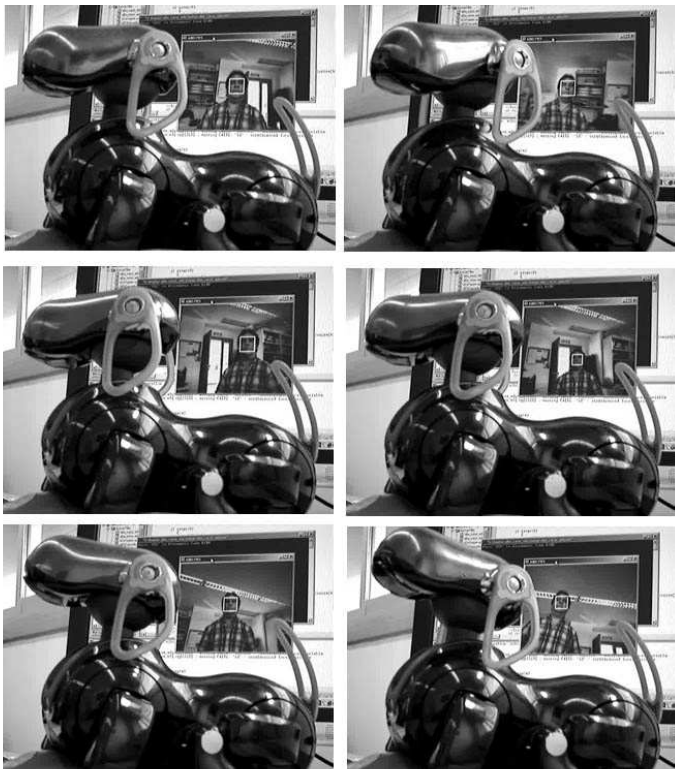


Figura 3. Detección y seguimiento de caras en tiempo real utilizando nuestro robot AIBO. El robot ajusta la posición/orientacíón de la cabeza en concordancia al movimiento de la persona. Se puede apreciar la robustez del algoritmo frente a los cambios de escala e iluminación.

investigador del programa Ramon y Cajal, del Ministerio de Educación y Ciencia de España.

Referencias

1. AIBO robot. http://www.sony.net/Products/aibo/index.html
2. Biederman, I., Mezzanotte, R.J., Rabinowitz, J.C.: Scene Perception: Detecting and Judging Objects Undergoing Relational Violations. Cognitive Psychology, **14** (1982) 143–177
3. Brézillon, P.: Context in Problem Solving: A Survey. The Knowledge Engineering Review, **14** (1999) 1-34
4. Brooks, R.A.: Intelligence without Reason. Proceedings of International Joint Conference on Artificial Intelligence (IJCAI), Sydney, Australia (1991) 569-595
5. Brooks, R.A, Stein, L.A.: Building Brains for Bodies. AI Memo No. 1439, MIT (1993)

6. Bruce, V., Young, A.: The Eye of the Beholder. Oxford University Press (1998)
7. Fischler, M.A., Elschlager, R.A.: The Representation and Matching of Pictorial Structures. IEEE Transactions on Computers, **COM-22** (1973) 67-92
8. Fodor, J. A.: The Modularity of Mind. MIT Press, Cambridge, Massachusetts (1983)
9. Kaplan, F., Oudeyer, P-Y.: Motivational Principles for Visual Know-How Development. Proceedings of the 3rd Epigenetic Robotics Workshop: Modeling Cognitive Development in Robotic Systems (eds. C.G. Prince, L. Berthouze, H. Kozima, D. Bullock, G. Stojanov and C. Balkenius), Lund University Cognitive Studies, Sweden (2003) 72-80
10. Kelly, K.: Out of Control. Addison-Wesley, New York (1994)
11. Minsky, M.: The Society of Mind. Simon & Schuster Publisher, New York (1988)
12. Newell, A., Simon, H. A.: Computer Science as Empirical Inquiry: Symbols and Search. Mind Design (ed. J. Haugeland), MIT Press, Cambridge, Massachusetts (1981) 35-66
13. Nilsson, N.J.: Introduction to Machine Learning. Draft book, Stanford University (1996) Available on Internet at: http://ai.stanford.edu/people/nilsson/mlbook.html
14. Ormrod, J.E.: Human Learning (3rd Edition). Merrill Prentice Hall, Upper Sadle River, New Jersey (1999)
15. Oudeyer, P-Y., Kaplan, F.: Intelligent Adaptive Curiosity: A Source of Self-Development. Proceedings of the 4th Epigenetic Robotic Workshop, Genoa, Italy (2004) pp. N/A
16. Shortliffe, E.H.: MYCIN: Computer-based Medical Consultations. Elsevier, New York (1976)
17. Singh, P.: The Open Mind Common Sense Project. MIT Media Lab. (2002) Available on Internet at: http://www.kurzweilai.net
18. Torralba, A., Sinha, P.: Statistical Context Priming for Object Detection. Proceedings of International Conference on Computer Vision, Vancouver, Canada (2001) 763-770
19. Torralba, A., Murphy, K.P., Freeman, W.T., Rubin, M.A.: Context-based Vision System for Place and Object Recognition. Proceedings of the International Conference on Computer Vision, Nice, France (2003) 273-280
20. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. International Journal of Computer Vision, **57** (2004) 137-154
21. Wada, K., Shibata, T., Saito, T., Sakamoto, K., Tanie, K.: Psychological and Social Effects of One Year Robot Assisted Activity on Elderly People at a Health Service Facility for the Aged. Proceedings of the International of the International Conference on Robotics and Automation (ICRA), Barcelona, Spain (2005) 2796-2801
22. Weiser, M.: The Computer for the Twenty-First Century. Scientific American **265** (1991) 94-104
23. Winograd, T.: Architectures for Context. Human Computer Interaction, **16** (2001) 401-419
24. Yokoyama, A.: The Possibility of the Psychiatric Treatment with a Robot as an Intervention - From the Viewpoint of Animal Therapy. Proceedings of Joint 1st International Conference on Soft Computing and Intelligent Systems and 3rd International Symposium on Advanced Intelligent Systems (2002) paper number 23Q1-1
25. Zajdel, W., Zivkovic, Z., Kröse, B.J.A.: Keeping Track of Humans: Have I Seen This Person Before?. Proceedings of International Conference on Robotics and Automation (ICRA), Barcelona, Spain (2005) 2093-2098

La inteligencia como propiedad física y la posibilidad de su explicación

Sergio Miguel Tomé

sergodel@terra.es

Resumen. Este artículo hace una revisión del estado actual de la inteligencia artificial como ciencia tras 50 años de existencia. El artículo pone de relieve el escaso desarrollo que ha mostrado la inteligencia artificial como ciencia, y el reducido conocimiento formal que se tiene de la inteligencia. La explicación de la inteligencia emerge como una de las principales cuestiones a las que se debe de enfrentar la inteligencia artificial para ser considerada una ciencia. En el artículo se propone considerar la inteligencia una propiedad física para intentar crear teorías matemáticas que la describan.

1 Introducción

La creación de máquinas inteligentes es un deseo que ha existido en el hombre desde las primeras civilizaciones. Los griegos, ya en su mitología, hablaban de máquinas con forma humana e inteligentes; aunque ha sido a partir del siglo XX, con la llegada de los computadores, cuando estas cuestiones han dejado de ser historias mitológicas para empezar a convertirse en una realidad. En el año 1956 nació la inteligencia artificial (I.A.), alumbrada en el "Darmouth summer research project on artificial intelligence" o más comúnmente conocido por la Conferencia de Darthmouth. El creador del término es John McCarthy, uno de los cuatro organizadores de la Conferencia de Darthmouth. Es interesante retroceder a la llamada que se realizó para convocar la conferencia de Darthmouth, ya que en ella se puede observar cuáles eran las intenciones de la nueva rama científica que se creaba. Si se lee la llamada se puede encontrar el siguiente párrafo:

> « La investigación será para proceder sobre la base de la conjetura de que cada aspecto del aprendizaje o cualquier otra característica de la inteligencia puede en principio ser tan precisamente descrita que una máquina puede ser fabricada para simularla.»

Aunque la I.A. es una disciplina realmente joven, la cual ahora tiene medio siglo, a priori, ha conseguido algunos éxitos notables. Tal vez, el hito más conocido de este joven campo científico haya sido crear una máquina que fue capaz de ganar al campeón del mundo de ajedrez; pero la realidad actual de la I.A. difiere de lo que en la conferencia de Darmouth se pretendía. ¿No debería la I.A. haber tratado de desarrollar una teoría que describiese minuciosamente las características de la inteligencia

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 46-56, 2006.

para desarrollarse como ciencia? Parece que la I.A. ha olvidado los objetivos que se proponían en su nacimiento y que se expresaban en el llamamiento de Darmouth. La realidad con la que nos encontramos es que la inteligencia artificial, como ciencia, apenas ha salido del cascarón. Acaso se podría hablar de la física como ciencia si la física no tuviera teorías que explicaran y describieran los fenómenos físicos.

El prometedor comienzo de la inteligencia artificial fue dando paso a dos líneas de investigación: una como ingeniería y otra como ciencia. La ingeniería, que se ocupa de crear programas y dispositivos con comportamientos útiles, ha sufrido un tremendo desarrollo. Pero la ciencia, por desgracia, ha permanecido prácticamente estancada en un rudimentario panorama para explicar lo que es la inteligencia. En mi opinión, podemos contemplar el estado de la inteligencia artificial como ciencia en la cronología que conforman los siguientes tres ítems. Primero, el temprano artículo en 1963 de Newell y Simon[11] "GPS, a program that simulates human thought" que, proponiendo una teoría algorítmica, GPS[10], sobre un determinado comportamiento en la resolución de problemas y su constatación experimental con sujetos humanos, hacía parecer que emergería rápidamente la inteligencia artificial como ciencia. Segundo, en los años 80 el estancamiento de la inteligencia artificial en su avance como ciencia puede observarse en importantes artículos como "Towards a General Theory of Action and Time"[1] o "A common representation for problem-solving and language comprehensión information"[3] en los que se intentan crear formalismos para la inteligencia artificial con un claro trasfondo de buscar la explicación de ciertos aspectos del pensamiento humano; pero a pesar de tener ese trasfondo se quedan en el campo de la algoritmia. Tercero, en 2003 Marvin Minky, gurú de la inteligencia artificial y uno de sus creadores, dijo públicamente en un discurso en la universidad de Boston "La inteligencia artificial padece de muerte cerebral desde los años 70".

La anterior cronología se origina por la confluencia de varias circunstancias. Una de ellas es que los iniciales éxitos fueron quedando atrás en el tiempo y los mecanismos que necesita una ciencia no aparecían, con el consiguiente enfriamiento del entusiasmo por crear una nueva ciencia que se ocupara de la inteligencia. Frente a ese panorama apareció otra circunstancia, el sector empresarial empezó a interesarse por la automatización de tareas debido a los beneficios económicos que podía y puede aportar. En el planteamiento del sector empresarial no interesa si existe una teoría que permita explicar la inteligencia, simplemente el dar los resultados y las respuestas correctas. No importa como surge un comportamiento sino que es el adecuado, eso es suficiente para que una solución sea correcta para el sector empresarial. Bajo ese panorama, la mayoría de los investigadores comenzaron a buscar cobijo en la algoritmia para encontrar soluciones a los problemas que el sector empresarial o la sociedad les planteaban. Así, una inmensa mayoría de los investigadores de inteligencia artificial siguieron ese camino, lo cual no puede ser en ningún caso depreciado, ya que gracias a ellos se puede hablar de la inteligencia artificial como ingeniería en toda regla y que no hay que dudar en apreciar.

Pero, ¿cuáles son los caminos que han seguido los investigadores de inteligencia artificial cuyo objetivo es explicar el fenómeno de la inteligencia de los seres vivos? Una de las principales líneas es la de las arquitecturas cognitivas. Hay varías arquitecturas cognitivas como: SOAR[6], ACT-R[2], COPYCAT[4], DUAL[5] o Subsumption Arquitectures [12]. A pesar de que las anteriores arquitecturas cognitivas tienen

sus bondades y sus logros, voy a realizar una crítica de ellas desde la perspectiva de tener un objetivo tan duro como explicar la inteligencia de los seres humanos. La clasificación no será por el parecido de sus técnicas sino por el principio general que las guía. Entre las arquitecturas cognitivas diferencio dos grandes grupos:

- **Arquitecturas que crean versiones algorítmicas de teorías psicológicas**

  Entre ellas, cabe destacar las arquitecturas: SOAR que tuvo a Newell como uno de sus creadores y Subsumption architectures que ha tenido como creadores a Alexandre Parodi y Rodney Brooks. A pesar, de las grandes diferencias entre ambas arquitecturas ambas intentan llevar a cabo la traducción de teorías de la psicología a la inteligencia artificial. SOAR es el intento de poner algorítmicamente las teorías cognitivas para ciertos aspectos de la mente posteriores a 1950. Actualmente, se está desarrollando la versión 9 de SOAR en la universidad de Michigan. A mi entender, Subsumption architecture es el intento de llevar a la inteligencia artificial la teoría conductista de la psicología; aunque si es verdad que el conductismo es un elemento dentro del comportamiento de los seres vivos, la psicología cognitiva puso de relieve demasiadas pruebas que apuntaban a que ciertos comportamientos son demasiado difícilmente explicables por medios del condicionamiento.

- **Las arquitecturas que proponen teorías algorítmicas para explicar propiedades cognitivas**

  Entre ellas se puede hablar de ACT-R, DUAL, o Copycat. La primera creada principalmente por John R. Anderson, la segunda por Boicho Coquimbo y la tercera por Douglas Hofstadter y Melanie Mitchell. A pesar, de sus grandes diferencias, creo que estas arquitecturas se plantean como meta la reproducción de comportamientos cognitivos. Así, intentan buscar esa reproducción básicamente desde el uso de paradigmas de programación, computación o técnicas matemáticas.

Desde mi opinión, las arquitecturas que crean versiones algorítmicas de teorías psicológicas no parecen unas candidatas a explicar totalmente la inteligencia por las razones que han sido mencionadas. Este tipo de arquitecturas tiene dos serios problemas. Primero están demasiado sujetas a las teorías que tiene la psicología, por lo que es difícil que puedan aportar explicaciones importantes que no estén en la propia psicología. Por otro lado, este tipo de arquitectura hereda un problema, y es que la psicología describe mayormente comportamientos externos. Así, cómo se sabe que es lo que verdaderamente pasa en un cerebro. Sería como ver a un ordenador al que se le da un vector de números desordenados y nos devuelve un vector con los números ordenados por orden creciente. Los observadores acuerdan que el comportamiento que realiza el programa que ejecuta es el de una ordenación; pero ¿qué algoritmo ha usado? Hay muchos algoritmos de ordenación, si nos pronunciamos por uno cabe la posibilidad de acertar, podemos equivocarnos un poco diciendo uno que use el mismo principio o equivocarnos totalmente al decir que algoritmo usa. Tal vez, midiendo tiempos de ordenación podemos acercarnos al tipo de algoritmo que usa el ordenador;

pero dar con el algoritmo exacto sólo sería una cuestión probabilística de acertar.

En el caso de arquitecturas que proponen teorías algorítmicas para explicar propiedades cognitivas su objetivo es reproducir comportamientos cognitivos y no procesos cognitivos, por lo que difícilmente lograrán explicar el fenómeno de la inteligencia. Al igual que en el caso de los algoritmos de ordenación, dos procesos cognitivos muy diferentes pueden dar como resultado un mismo comportamiento cognitivo.

Además de los problemas expuestos, muchas de estas arquitecturas cognitivas han desembocado en la creación de entornos de desarrollo de software que tienen como herramientas la versión algorítmica de ciertas teorías psicológicas sobre la mente. Así, la imagen de una ciencia con su teoría matemática para describir el fenómeno de la inteligencia queda muy lejos de la realidad del campo de la inteligencia artificial. Así, a pesar de que la inteligencia artificial se haya desarrollado potentemente como ingeniería, no se puede ocultar que como ciencia sus objetivos no se han cumplido y se encuentra en una situación rudimentaria.

2 La inteligencia como propiedad física de la naturaleza

Actualmente, la inteligencia se considera un fenómeno psicológico o un fenómeno biológico. Existe una rivalidad entre los que opinan que la psicología y sociología son parte de la etología y, por tanto, de la biología, y los que opinan que no son parte. Aunque, el desarrollo en los últimos años de la neuroetología ha hecho ganar adeptos a considerar la inteligencia un fenómeno biológico. La biología es la rama de la ciencia encargada del estudio de la vida, cómo las especies logran sobrevivir y sus interacciones con otros organismos y el entorno. Por ello, la concierne las características, la clasificación y el comportamiento de los organismos.

La primera propuesta que pretende lanzar este artículo es la de que la I.A. debería considerar que la inteligencia es un fenómeno físico. La consideración de que la inteligencia es un fenómeno biológico o psicológico no es el mejor punto de vista para la inteligencia artificial y su desarrollo como ciencia. Antes de entrar en los detalles de si se puede considerar legítima la propuesta, el lector se estará preguntado por qué se quiere realizar este cambio de consideración. Las razones para pretender dar este cambio son dos:

Primero, si la inteligencia artificial tiene como objetivo lograr que una computadora exhiba inteligencia total(objetivo de Dartmouth), ya no se puede decir que la inteligencia sea un fenómeno biológico. Así, no tiene mucho sentido para un investigador de I.A. el considerar que la inteligencia es un fenómeno biológico (salvo que existiera una demostración de que la conjetura de Dartmouth es falsa, cosa que por ahora no ha ocurrido); ya que si el objetivo se completara habría que cambiar la propia definición de vida. Si alguien piensa en salvar ese problema alegando que sólo se intenta simular la inteligencia, se equivoca en su argumento, porque si se pretende que una computador tenga inteligencia total, se está diciendo que sea indistinguible de la inteligencia. De modo que, si es indistinguible es que es inteligencia. No se simula que se juega al ajedrez en un ordenador, se juega al ajedrez en un ordenador.

Segundo, la biología o la psicología no suelen describir los sistemas en términos de

objetos que obedezcan leyes que sean descritas matemáticamente. Si se considerase la inteligencia una propiedad física sería ineludible la creación de teorías matemáticas que la describieran. Eso ayudaría enormemente al investigador de I.A., dado que el computador es la herramienta fundamental para la I.A. y la conexión entre matemáticas y computadores es más que notable. De otro modo, se obliga a pasar las teorías que se tengan sobre la inteligencia a formalismos matemáticos que puedan ser implementados en un computador. Por ejemplo, recuérdese la arquitectura cognitiva SOAR.

Por las razones expuestas, se quiere proponer que dentro de la I.A. se haga la consideración de que la inteligencia es una propiedad física. Ahora, el lector se estará preguntando, ¿a qué se refiere la palabra propiedad? La idea es considerar la inteligencia como una propiedad física, con las mismas consideraciones que hace la física sobre las propiedades físicas ya aceptadas por todos. Para entender la argumentación a favor de considerar la inteligencia una propiedad física se fijarán dos términos usados en física: propiedad física y fenómeno físico.

- Propiedad física. Una propiedad física de un sistema (objeto real más o menos complejo y dotado de algún tipo de organización) es una característica que puede ser estudiada usando los sentidos o algún instrumento específico de medida.
- Fenómeno físico. Un fenómeno físico es cualquier proceso natural entre sistemas que es observable, posible de ser medido, y donde no hay transformación[1] de la materia que compone los sistemas que intervienen.

Entre los dos conceptos anteriores existe una dualidad, ya que toda propiedad física lleva asociada un fenómeno físico y viceversa. Esta dualidad es obligatoria, ya que si una propiedad física no llevara asociada un fenómeno físico que proporcione una interacción de la materia con su entorno no habría ningún modo de saber que la materia tiene esa propiedad física. Por ejemplo: la propiedad física de la temperatura y el fenómeno físico del calor; imagínese que se quiere medir la temperatura de un radiador. Para ello se coge un termómetro y se acerca al radiador. Entonces mediante el fenómeno físico del calor se transfiere energía al termómetro, de manera que hace que el termómetro adquiere energía y cambie su temperatura. El fenómeno físico del calor se produce sólo cuando hay una diferencia entre la temperatura de dos sistemas, así que cuando los dos alcancen la misma temperatura el fenómeno físico parará y se sabrá cual es la temperatura del radiador porque será la misma que la del termómetro. Si no existiera la dualidad de la propiedad física y fenómeno físico, no se podría acceder a medir la propiedad física del sistema. Otros ejemplos de propiedad y fenómeno son: masa, campo gravitatorio; carga eléctrica, campo eléctrico; densidad, flotabilidad;...

Posiblemente, al principio, pueda resultar chocante la propuesta que se está realizando. Así, para justificar la legitimidad de la propuesta se mostrará que según la definición y comparando con otros fenómenos físicos, aceptados por todos, la propuesta está justificada. La justificación se llevará a cabo estableciendo un paralelismo

[1] Si hubiera transformación de la materia sería un fenómeno químico.

entre un fenómeno físico admitido por todos y la propuesta de contemplar la inteligencia como un fenómeno físico. Pero antes de continuar, se debe dar una mínima definición de qué se entiende en este artículo por la propiedad física de la inteligencia y el fenómeno físico que lleva asociado.

> « La inteligencia es la propiedad física por la que una entidad es capaz de modificar el estado de la realidad dirigiéndolo hacia una situación específica, a condición de que la modificación no se produzca de manera aleatoria. »

> « El comportamiento inteligente es el fenómeno físico por el que una entidad modifica el estado de la realidad dirigiéndolo hacia una situación específica, a condición de que la modificación no se produzca de manera aleatoria. »

Estas definiciones las justifico en la teoría del Profesor Rodolfo Llinás[7] sobre la aparición del cerebro en los seres vivos. El profesor Llinás sugiere que el cerebro surge para cubrir la necesidad de realizar movimientos que dirijan a la entidad hacia una situación específica. Esa necesidad sería debida a que los animales tienen que moverse en el mundo externo y, por lo tanto, requieren una imagen, aunque sea muy primitiva, de hacia dónde se están moviendo, ya que sin ella podrían estar dirigiéndose hacia la boca del depredador. Sus movimientos deben de llevarles hacia una situación en la que ellos puedan comer y evitarles las situaciones en las que puedan ser comidos. Un movimiento aleatorio no conlleva ninguna ventaja evolutiva. El Profesor Llinás basa su respuesta en pruebas de la biología como los tunicados. Los tunicados son seres que viven en el fondo del mar, son como una especie de botella, sólo toman agua y la empujan con un filtro. Este sistema tan mínimo no requiere cerebro; pero cuando estos animales se reproducen, generan una semilla que contiene un cerebro. La semilla de los tunicados que es móvil como un renacuajo tiene capacidad de recibir luz, sabe dónde es arriba y dónde es abajo; tiene la posibilidad de entender muy brevemente el mundo externo. Este ser se mueve activamente y busca un sitio donde fijarse. Cuando el tunicado encuentra ese lugar se fija en él, mete la cabeza y absorbe su propio cerebro como un nutriente más. La razón de este comportamiento es que ya no lo necesita, porque ya no se moverá más.

Otra aclaración que se debe realizar es que en este artículo la inteligencia se considera algo diferente de la percepción a pesar de su innegable dependencia. En ningún momento, cuando se habla de inteligencia en este artículo se usa de un modo que contenga a la percepción.

2.1 Justificación de la propiedad física de la inteligencia

Como anteriormente se mencionó se va a realizar un paralelismo con una propiedad física y un fenómeno físico aceptados por todos y la inteligencia como propiedad física y comportamiento inteligente como su fenómeno físico asociado. La propiedad física que se va a usar es la temperatura y su propiedad física el calor, que se definen

a continuación.

- El calor es el fenómeno físico por el que se transfiere parte de la energía interna de un sistema a otro, con la condición de que estén a diferente temperatura.
- La temperatura es la propiedad que mide la cantidad de energía cinética contenida en un sistema y asociada al movimiento aleatorio de las partículas que lo componen.

Ahora, procedamos a realizar una comparativa entre la observabilidad y medibilidad de ambos fenómenos y propiedades físicas:

Calor y temperatura: Se pueden observar al poner en contacto un termómetro con una bola de metal que estén a diferente temperatura. En el momento en que pongamos en contacto el termómetro con la bola de metal veremos como el mercurio del termómetro empieza a dilatarse hasta que los dos sistemas alcancen un equilibrio térmico. Gracias, a que el coeficiente de dilatación del mercurio permanece aproximadamente constante, el termómetro nos sirve para medir la temperatura de cualquier sistema una vez establecida una temperatura de referencia y su correspondiente escala. Una unidad para medir la temperatura es en grados centígrados, y para el calor las calorías

Comportamiento inteligente e inteligencia: Pongamos un niño de 4 años al que le gustan las golosinas en una habitación con una silla y una estantería con una bolsa de golosinas visible; pero colocada a una altura que el niño no alcanza de manera natural. Al niño le está permitido coger las golosinas. En esa situación, emergerá un comportamiento en el que el niño cogerá la silla, se subirá a ella, alcanzará las golosinas y empezará a ingerirlas. Es evidente que el fenómeno del comportamiento inteligente es observable. Además, esto no es ninguna novedad, la observación del comportamiento inteligente y la medición de la inteligencia son cosas que la psicología lleva realizando mucho tiempo y que no puede sorprender a nadie. El método consiste en hacer que una entidad se enfrente a una batería de problemas. Si la entidad tiene como propiedad la inteligencia, en el entorno creado por la batería de problemas y la entidad, se dará el fenómeno del comportamiento inteligente de manera que se obtendrá un número modificaciones en el estado del entono. Estas modificaciones del entorno se miden en soluciones correctas realizadas. Mediante el número de soluciones correctas y otros datos se puede calcular el cociente de inteligencia.

En ambos fenómenos físicos no se produce una transformación de la materia. En el primer ejemplo el mercurio del termómetro sigue siendo mercurio, el cristal sigue siendo cristal y la bola de metal sigue siendo del mismo metal. En el segundo caso el niño sigue siendo el mismo niño, la silla sigue siendo la misma silla y los caramelos siguen siendo los mismos caramelos (porque una vez ingeridos hay transformación de la materia que son debidas a fenómenos químicos que se dan en el interior de cuerpo del niño, pero que no forman parte del fenómeno de comportamiento inteligente).

A pesar de la similitud con propiedades físicas aceptadas por todos, el lector puede estar considerando que existe una diferencia importante, la estructura en la que esté organizado el cerebro da como consecuencia que la propiedad de inteligencia de unos valores u otros, cosa que no ocurre con la temperatura. Es cierto, pero también hay propiedades y efectos físicos con los que ocurre lo mismo. Pongamos los dipolos magnéticos y el magnetismo. La intensidad del fenómeno del magnetismo de un sistema depende de la organización de los dipolos magnéticos del sistema, el término de la física para reflejar que existe una dependencia de la organización de los dipolos magnéticos es el de dominio magnético. No hay que tener reticencia porque la organización del sistema influya en el valor neto de la propiedad y del fenómeno, ya que en la física hay propiedades y fenómenos que dependen de la configuración del sistema y no por ello dejan pierden su consideración de propiedad y fenómeno. Después de estás comparaciones, creo que, al menos, queda justificada la postura de considerar la inteligencia una propiedad física.

2.2 La inteligencia como propiedad matematizable

Si se acepta la propuesta de entender la inteligencia como una propiedad física y el comportamiento inteligente como un fenómeno físico, el primer paso de la inteligencia artificial debe ser comenzar un serio programa de investigación con el objetivo de obtener una teoría matemática sobre la propiedad de la inteligencia que describa y explique la propiedad de la inteligencia y su fenómeno asociado.

El primer paso de la obtención de una teoría matemática de la inteligencia sería encontrar un marco matemático adecuado para describir la propiedad y el fenómeno físico. La física describe diferentes fenómenos y propiedades de la naturaleza, y para cada uno de ellos usa el formalismo más adecuado: para el espacio se utiliza la geometría, para la temperatura las ecuaciones diferenciales y para la mecánica clásica mediante el cálculo vectorial o de variaciones. Pero, ¿cuál es la rama de las matemáticas que se puede usar para describir el fenómeno de la inteligencia? Para responder a está pregunta, recuérdese que la definición que se daba de la propiedad de inteligencia aludía expresamente a la capacidad de modificar un entorno hacia un estado definido. Parece lógico que el marco matemático debe de poder describir entornos y las modificaciones que se pueden realizar sobre este. Como ya se mencionó en este artículo, no se considera el concepto de percepción dentro de la propiedad de la inteligencia. Esto quiere decir, que una teoría matemática que se construya para describir y explicar la propiedad de la inteligencia no tiene que explicar como un sistema puede ser capaz de percibir el entorno, sus objetos y propiedades. La teoría deberá de intentar explicar como dada una percepción del entorno debido a la propiedad de la inteligencia el sistema adoptará un determinado comportamiento, y le deberán ser transparentes los mecanismos y procesos que ocurren para percibir un entorno. Bajo ésta consideración, el marco matemático más adecuado para desarrollar una teoría sobre el fenómeno de la inteligencia es la teoría de modelos, ya que trata de estructuras y sus descripciones mediante lenguajes, siendo en este marco matemático transparentes las cuestiones de la percepción.

La teoría de modelos [8] es una rama de la lógica-matemática que se ocupa de describir las estructuras matemáticas. La teoría clásica de modelos proviene de los años 50's del siglo XX, pero sus antecedentes son de varias décadas atrás. El impulsor de la teoría de modelos fue el lógico polaco Alfred Tarski, a él cabe la concepción y dirección de un programa de investigación en la Teoría de Modelos. En 1934, Tarski dió con precisión la definición de los conceptos absolutos de satisfacción, verdad y consecuencia, pero no fue hasta 1957 cuando junto a Vaught definió satisfacción y verdad en un sistema. Desde 1947 Tarski enseño en el departamento de matemáticas de Berkeley y a su alrededor se formó un grupo de investigadores que en los años 50's produjo la teoría clásica de modelos. En la teoría de modelos se distinguen dos tipos de realidades, las estructuras matemáticas y los lenguajes fórmales, y estudia las relaciones que hay entre los dos tipos de realidades. La semántica se ocupa de conectar estos dos tipos de realidades mediante la noción de verdad. Tarski [13] solía dar como ejemplo del tipo de conexión que se crea la siguiente frase:

« "La nieve es blanca" es verdad si, y sólo si, la nieve es blanca.»

"La nieve es blanca" es una fórmula del lenguaje. Esta fórmula toma su significado (o su valor de verdad) cuando está conectada con una realidad. Así, la oración es verdad, cuando sobre la realidad que se hace la afirmación se cumple la afirmación. Es decir, la fórmula "la nieve es blanca" es verdad si se cumple el hecho físico de que la nieve es blanca.

2.3 Requisitos para una teoría matemática de la inteligencia en el marco de la teoría de modelos

La creación de una teoría sobre la propiedad de la inteligencia en el marco de la teoría de modelos debe tener como primer requisito la construcción de una clase de estructuras matemáticas que pueda representar la realidad que rodea a los seres vivos y las características que estos perciben. Así, la estructura matemática será biyectable con la realidad; aunque no se trata de hacer leyes de física, sino que la estructura refleje en sus elementos lo que el ser vivo puede percibir en la realidad. Si se quiere crear una estructura para una teoría sobre la inteligencia que tienen los seres humanos es necesaria una estructura matemática con las siguientes características:

- **Pasado, Presente Futuro**: Es necesario que exista una perspectiva temporal del entorno.
- **Estados hipotéticos y existentes de la realidad:** Los seres humanos no sólo son capaces de pensar en el estado de realidad que están percibiendo, también pueden pensar los estados en que pudiera estar la realidad.
- **Estado del entorno**: Necesita representarse cada estado posible de la realidad.
- **Momentos del tiempo:** Debe indicar los distintos momentos del tiempo.
- **Acciones:** Las acciones que realizan los objetos hacen evolucionar al estado de una realidad.

Dentro del desarrollo de una teoría de la inteligencia la estructura matemática cumplirá la función de describir el fenómeno del comportamiento inteligente, ya que es el fenómeno del comportamiento inteligente el efecto que se produce en una realidad debido a la propiedad de la inteligencia que tenga un sistema.

El segundo paso en el desarrollo de una teoría matemática de la inteligencia es estudiar los lenguajes formales para la descripción de la estructura matemática desarrollada en el primer paso. Estos lenguajes son la base sobre la que se debe explicar la propiedad de la inteligencia. Sobre ellos se debe explicar la capacidad de modificar el estado de la realidad dirigiéndolo hacia una situación específica.

En ese proceso de crear una teoría matemática sobre la propiedad de la inteligencia y el fenómeno del comportamiento inteligente en el marco de la teoría de modelos será vital encontrar una relación que ligue la propiedad al fenómeno. Trasladado al marco de la teoría de modelos, se habrá de encontrar algún elemento del ámbito de los lenguajes formales que imponga una restricción a las estructuras matemáticas. De manera, que la teoría sobre la inteligencia permita encontrar el conjunto de estructuras en las que exista la dualidad de propiedad y fenómeno en cada una de ellas.

3 Conclusiones

En este artículo nos hemos interesado por el panorama actual de la inteligencia artificial como ciencia tras cumplirse cincuenta años de su creación. El resultado de la indagación ha mostrado un estado realmente rudimentario. Pero a pesar del estado actual de la inteligencia artificial, este artículo ha intentado mostrar que la inteligencia artificial no está avocada a ser una mera traducción de otras ciencias, como la psicología, sino que puede ser una ciencia activa en la investigación de la inteligencia; aunque para ello, como propone el artículo, será necesario un drástico cambio del punto de vista de la comunidad científica que trabaja en la inteligencia artificial.

Aunque el artículo, se ha ocupado principalmente de mostrar una línea para la creación de una teoría sobre la inteligencia que haga de la inteligencia artificial una ciencia, hay cuestiones que no se han mencionado, pero que habrá que tener en cuenta si la inteligencia artificial quiere desarrollarse también como ciencia. Una de esas cuestiones es la realización de experimentos controlados que propugna el método científico para comprobar una teoría. Pero no se puede querer aplicar el método científico a la inteligencia artificial como a la física, que es el prototipo de ciencia. De esto ya hablo Newell [9] exponiendo el problema del establecimiento de condiciones iniciales para llevar a cabo comprobaciones de una teoría sobre la inteligencia.

Referencias

1. ALLEN, J.F. (1984): "Towards a General Theory of Action and Time" En Artificial Intelligence 1984; 23 Págs. 123-154.
2. ANDERSON, J. R. (1996): "ACT: A simple theory of complex cognition" En American Psychologist, 51, 355-365.
3. CHARNIAK, E.(1981)"A common representation for problem-solving and language com-

prehensión information”. En Artificial Intelligence 1981; 16 (3) Págs. 225-255
4. HOFSTADTER, D., & MITCHELL, M. (1994): “The Copycat project: A model of mental fluidity and analogy-making”. En Holyoak, K. and Barnden, J. (Eds.), Advances in Connectionist and Neural Computation Theory, Volume 2: Analogical Connections (pp. 31-112). Ablex.
5. KOKINOV, B. (1994): “The DUAL cognitive architecture: A hybrid multi-agent approach” En A. Cohn (Ed.), Procedente de la Eleventh European Conference on Artificial Intelligence. London: John Wiley & Sons, Ltd.
6. LAIRD, J. & NEWELL, A. & ROSENBLOOM, P. (1987). "Soar: An Architecture for General Intelligence". En *Artificial Intelligence*, 33: 1-64.
7. LLINÁS, R.(2001):I of the vortex, Cambrige, MA,MIT press
8. MANZANO, M. (1989): Teoría de modelos, Madrid, Alianza Editorial (Existe una versión publicada por Oxford Press).
9. NEWELL, A. (1981, Summer): The Knowledge Level. En AI Magazine, 1. Págs. 1-20.
10. NEWELL, A. SHAW, J.C. SIMON, H.A. (1960) “Report on a general problem solving program for a computer” Information processing: proc. Of the International Conference on Information Processing. UNESCO, París, pp. 256-264.
11. NEWELL, A. SIMON, H.A. (1963) “GPS, a program that simulates human thought ” En CT. 279, 293
12. NITAO J. J. & PARODI A. M. (1986) "A Real-Time Reflexive Pilot for an Autonomous Land Vehicle". En IEEE Control Systems, vol. 6, no. 1, February, pp.14-23.
13. TARSKI, A. (1944): ``The semantic of Truth and Foundations of Semantics". En Philosophy and Phenomeno logical Research Págs. 341-374

Esbozo de una lógica del ver: Fundamentos, método y conexiones

Eduardo Álvarez Mosquera

eduardoalvar@gmail.com

Resumen. Esta comunicación tiene varias puntas. Primero que nada, plantea la posibilidad de construir una lógica con base empírica con un claro objetivo, legitimar ciertas prácticas de ver humanas. Segundo, tal lógica supone un trabajo, por decirlo así, de equipamiento. Hay que dotarla de una batería de reglas, también empíricas, que atiendan a lo que cada hombre hace cuando ve o lo que legítimamente puede afirmar cuando ve. Y eso es lo que se hace. Tercero y último, que partiendo de los resultados obtenidos, es posible pensarla como una ayudante de la psicología. Más aún, como la proveedora de material para un proyecto de trabajo conjunto. Y ¿en qué queda todo eso?. En una invitación para formar parte de un plan de mayor aliento, la tercera cultura.

1 La cuestión

En cuanto leí el programa del congreso y las exigencias para enviar "comunicaciones", no pude dejar de sentir sorpresa y agrado. Se prometía que no iba a ser un club exclusivo de científicos como quería Brockman. En cada panel se anunciaba ciencia, en cada panel se anunciaba filosofía. Era como asistir a la esperada "tercera cultura" de Snow. Y la verdad, no quería perdérmelo.

No obstante, había una pregunta que me inquietaba: ¿con qué podría contribuir aquí la filosofía?. Y lo que mejor me pareció, parafraseando a Habermas, era hacerla funcionar como mediadora (1), mediadora entre la lógica y la psicología. El cómo, formulando una lógica del ver que pudiera articularse con la percepción natural y la psicoterapia. Por esto mismo separé esta comunicación en tres partes. En la primera parte quedará establecida la acepción del vocablo ‘ver’ y lo que a partir de él puede ser dicho. La segunda parte se destinará a la exposición de las características generales de esta lógica del ver, de sus reglas, y además, de lo que en ella puede deducirse. Y al final cómo es que a partir de esta lógica se puede llegar a pensar en un cambio de dirección en la manera de concebir a la psicología y a la terapia psicológica.

2 Ver y decir

Para comenzar vayamos reconociendo un hecho, la anfibología del vocablo ‘ver’. Lamentablemente ‘ver’ se puede emplear en muchos sentidos; luego, estoy obligado a fijar el sentido en que voy a usarlo.

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 57-64, 2006.

Primera nota: el ver es aquí el ver humano; otros modos de ver, por ejemplo el de los distintos animales, no se considerarán pertinentes. Aquí solo están implicados el cuerpo humano, lo que le pasa a él y lo que hace interiormente con lo que le viene de fuera. Segunda nota: el ver siempre es dependiente de una situación objetiva. En este punto sigo a Ortega y Gasset y a su célebre ejemplo del paisaje. Él cree que dado un paisaje, según sea la situación del que ve, ese paisaje se organiza de una manera o de otra. Por esto mismo, para dos hombres en distinta situación, el paisaje de uno diferirá del paisaje del otro, y aunque no cambie, lo que para el primero es nítido, para el otro puede ser oscuro o borroso (2). Tercera nota: ver es ver desde cierta cultura. Sin duda, hay en el ver un fuerte componente no natural que supone un proceso de socialización del individuo que ve. En otras palabras, se aprende a ver.

En esto sin duda, ya puede entreverse un triángulo formado por el hombre, los objetos y la sociedad, en el cual está implicado un toma y daca entre lo biológico y lo social. Pero ese triángulo no nos interesa, nuestro asunto no es la génesis del ver. Por eso es que la cuarta nota es: el ver está directamente relacionado con el conocimiento del que ve. De esta manera, el ver se constituye en un elemento central capaz de definir y permitirnos hablar sobre cualquier tipo de objetos (3).

Quedémonos con esta última idea y examinemos algo bien sencillo. Me refiero a la expresión '**a** ve a **b**', siendo **a** y **b** la abreviatura del nombre de dos personas diferentes. Supongamos ahora que el que dice '**a** ve a **b**' es **a**. Bajo esta hipótesis, lo más razonable es pensar que **a** no está hablando por hablar y que tiene buenos argumentos para sostener eso. Si esto es cierto entonces, la cosa sería como dicen Russell (4) y Hintikka (5):

Veo P1, Veo P2, Veo P3, ..., Veo Pn; luego, de hecho veo a **b** (6).

En otras palabras, **a** estaría viendo P1, P2, P3, ..., Pn, que no son otra cosa que todas las propiedades, compatibles entre sí, que **a** le asigna a **b**. A partir de ahí, **a** se cree con derecho a decir que ve a **b**.

Ahora bien, ¿le asiste realmente ese derecho a **a**?. Si pensáramos en que esa inferencia debe ser consciente, al menos en casi todos los casos la respuesta tiene que ser no. En realidad, que las cosas ocurriesen de esa manera sería muy raro. En cambio, si pensáramos con Russell, que lo habitual es hacer inferencias sin tener conciencia de que las hacemos, la respuesta sería sí. Me adscribo a esta posición.

Mantengamos la expresión '**a** ve a **b**', pero cambiemos al que la profiere. Si no es **a** el que habla, tendrá que ser o bien **b** o bien cualquier otra persona cuyo nombre abreviaré con **c**. Sin duda esto no es indiferente. Si el que habla es **b**, se estaría en el caso de ver que se es visto, y si el que habla es **c**, en el de ver que alguien ve. Para decirlo con un ejemplo: para una mujer no es lo mismo ser vista por un varón, que ver que ese varón ve a otra mujer.

No obstante esto, ambos casos tienen algo en común. Me refiero a que es un ver que otro ve, para el cual no faltan argumentos. Aquí sostengo que se realiza una inferencia del tipo:

Veo PV1,Veo PV2, Veo PV3, ...,Veo PVn, luego, de hecho veo el ver de **a**.

La diferencia con el esquema inferencial anterior es evidente. Difieren las propiedades, no en cuanto a propiedades, sino en cuanto que ahora son propiedades compatibles que se asignan al modo de ver. Se ve que **a** hace tal cosa, tal otra, etc., y como hace todas esas cosas, se tiene derecho a decir que '**a** ve a **b**'.

Pero demos un paso más y sustituyamos ahora a **b** por **a**. Así, la expresión '**a** ve a **b**' quedaría '**a** ve a **a**'. Es una nueva expresión y tiene la particularidad de reflejar una experiencia harto común, la de verse a sí mismo. Todos los días me veo en el espejo; este es un ejemplo claro.

Más, esto no es todo lo que veo; también veo a mi esposa viéndose frente al espejo. Este sería un caso de '**b** ve a **b**' dicho por **a**. La cuestión es entonces, saber si tanto yo como mi esposa tenemos derecho, o no, a decir que nos vemos en el espejo y si yo tengo derecho, o no, a decir que la veo verse en el espejo. Para el primer caso, en donde **a** sostiene que '**a** ve a **a**' por decir lo mínimo, no parece haber dificultades. Cambiando a **b** por **a**, es posible entenderlo según el primer esquema de inferencia. Para el segundo caso, en el que **a** ve que '**b** ve a **b**', no habría reparo alguno tampoco. Lo único que cambiaría sería que ahora el esquema de inferencia es el segundo y que habría que sustituir a **a** por **b**.

Quedémonos aquí y con esto, ya hemos aprendido algunas cosas. Lo primero, que el ver es siempre un ver fundamentado. Ver una cosa no es un simple dato; antes que nada es el resultado de una inferencia. Segundo, que a pesar de que el ver es fundado, esto no significa que sea concluyente. Es posible pensar en que para diferentes personas, por más que sigan obligatoriamente el mismo esquema inferencial, lo que vean cada una de ellas sea muy distinto. En tercer lugar, que el ver conecta a cada cual con un modo de pensar la realidad y hablar sobre ella. Por el ver afirmo que tales y cuales cosas existen, son de determinada manera, que me agradan o no las tolero más, etc.. En otras palabras, que por el ver estoy legitimado en pensar lo que pienso y en decir lo que digo.

3 Lógica del ver

3.1. Fundamentos y método

Lo básico está ahora encima de la mesa. La cuestión siguiente es explicar cómo es que es posible una lógica del ver. En torno a esto tengo que decir dos cosas que no son del todo ortodoxas. La primera de ellas, que la idea de una lógica del ver me vino a raíz de una obra literaria de Sartre, "La infancia de un jefe", obra de la cual tendremos alguna noticia más adelante. La segunda, que la lógica del ver es una especie de cocktail. Al principio es una idea que asusta, pero en cuanto reparamos en que esto mismo ha sido dicho por Garrido del ilustre silogismo (7), dejamos de preocuparnos por eso. Lo que cuenta aquí es que en ese cocktail, los ingredientes son:

- una lógica al estilo de Nagel. Con eso quiero decir una lógica naturalista, y por ende, desprovista lo más posible de metafísica.
- la lógica proposicional presentada bajo la forma de deducción natural a la manera de Gentzen, con reglas incluidas.

- nuevas reglas de inferencia, que serían específicas de esta lógica del ver.

Señalado esto, lo primero que queda claro es que la lógica proposicional va a estar incluida en la lógica del ver. De ella tomará sus verdades apriórícas y hasta su método. Luego, debe considerársela un auxiliar imprescindible.

No obstante, la lógica del ver diferirá en mucho de la proposicional. De ningún modo aspira a reglar el ver humano, su asunto no es cómo debieran ver los hombres. Más bien es lo contrario. Tiene como objeto cómo es que de hecho ven los hombres y lo evalúa. Esto quiere decir básicamente tres cosas. Que en cuanto todo ver es un ver de tal o cual persona, a la lógica del ver no le queda más remedio que aceptar ser ciencia de lo particular (a). Además, que tienen que quedar fuera como no pertinentes predicaciones lógicas del tipo es 'correcto' o 'incorrecto'. En ella se dirá 'tiene derecho a decir' o 'no tiene derecho a decir' (b). Y finalmente se tendrá que reconocer como una lógica sin fin. Ha de admitir la posibilidad de modos de ver emergentes, y por lo tanto, de ir incrementando a lo largo del tiempo sus reglas y sus demostraciones (c).

Queda sellado así el status de la lógica del ver. Sería una ciencia con base empírica que busca legitimar la experiencia de ver a través de pruebas formales, y en la cual lo posible solo entra, no como un ver mundos posibles, sino como posibles modos de ver este mundo. Dicho en pocas palabras, se constituiría en una lógica de este, nuestro mundo.

3.2. Algunas reglas y algunos problemas de la lógica del ver

Ahora el asunto es poner todas estas ideas a marchar. Comencemos antes que nada con el vocabulario que vamos a utilizar. Dispondremos de tres tipos de símbolos: símbolos lógicos, símbolos no lógicos y símbolos auxiliares.

Los símbolos lógicos son conectores: -, ., v, →, ↔. Los símbolos no lógicos, en cambio, son de distinto orden. Tenemos un operador: V, un definidor: =, letras predicativas: P, Q, R, ..., P1, Q1, R1, ..., Pn, Qn, Rn y letras individuales. Estas últimas se dividen en dos grupos: las letras individuales inespecificadas: p, q, r, ..., p1, q1, r1, ..., pn, qn, rn, y letras individuales especificadas: a, b, c, ..., a1, b1, c1, ..., an, bn, cn. Finalmente están los símbolos auxiliares, que no son otra cosa que símbolos separadores: ( ), (),{ }. Expliquemos esto. Con respecto a los símbolos lógicos digamos que '-' es igual a 'no', que '.' es igual a 'y', que 'v' es igual a 'o', que '→' es igual a 'si ... entonces', y que '↔' es igual a 'si y solo si'.

Los símbolos no lógicos son un poco más complicados. Tenemos el operador 'V', con el cual se simboliza 've a' y el definidor '=', que representa a 'es igual a'. Por otra parte están las letras. Las predicativas, P, Q, R, P1, Q1, R1, ..., Pn, Qn, Rn, sirven para indicar propiedades asignables a personas o cosas. Las individuales finalmente, cuando son inespecificadas indican que en su lugar puede escribirse cualquier letra individual especificada, y cuando son especificadas están representando al nombre de una persona o cosa. Los símbolos auxiliares, quizá los más modestos pero no por eso menos importantes, tienen como misión separar.

Una vez sabido esto, pasemos a las reglas. Por cierto, aquí no estarán todas, solo enunciaré unas pocas; las que necesito para poder trabajar con los problemas que voy

a considerar aquí. Hablaré de tres tipos de reglas: reglas de sustitución (RS), reglas de reducción (RR) y reglas de especificación (RE).

Reglas RS. Aquí me referiré a una sola, a aquella que se aplica en el comienzo y en el final de la deducción y a la que denominaré RS1. Se usa para sustituir a una letra individual inespecificada por una letra individual especificada y viceversa (8).

Versión 1: RS1-1 p/a Versión 2: RS1-2 a/p

RS1-1 se lee: dada una expresión en la que figure la letra p, se está autorizado a sustituirla por la letra a.

RS1-2 se lee: dada una expresión en la que figure la letra a, se está autorizado a sustituirla por la letra p.

Reglas RR. De un modo general, son reglas que permiten pasar de una expresión a solo una parte de ella. Con su aplicación se puede eliminar el operador V. Ejemplos de reglas RR:

RR1: autoriza a eliminar V, independientemente de si soy yo u otro quien afirma que ve.

Versión 1: RR1-1 Vp Vq r ├ Vq r Versión 2: RR1-2 Vr Vq r ├ Vq r

RR1-1 se lee: dada una expresión con las letras p, q y r, y siendo el caso que p ve que q ve a r, entonces se está autorizado a pasar a: q ve a r.

RR1-2 se lee: dada una expresión con las letras p, q y r, y siendo el caso de que r ve que q ve a r, se está autorizado a pasar a: q ve a r.

RR2: Esta es una variante de la regla anterior. Al igual que RR1, autoriza a eliminar a V, pero ahora en una expresión que contiene el definidor =.

Notación: RR2 Vp q=P ├ q=P.

Se lee: dada una expresión con las letras p, q y P, y siendo el caso de que p ve que q es igual a P, se está autorizado a pasar a: q es igual a P.

Reglas RE=. Estas son reglas que permiten especificar lo que está implicado cuando se ve. Aquí se va a enunciar la regla RE=1, que es la que autoriza a conectar el ver con el definir.

Versión 1: RE=1-1 Vp q ├ Vp q=P Versión 2: RE=1-2 Vp q=P ├ Vp q

RE=1-1 se lee: dada una expresión con las letras p, q y P, y siendo el caso de que p ve a q, entonces se está autorizado a pasar a: p ve que q es igual a P.

RE=1-2 se lee: da una expresión con las letras p, q y P, y siendo el caso de que p ve que q es igual a P, entonces se está autorizado a pasar a: p ve a q.

Final de las reglas, y comienzo del trabajo deductivo. Pasemos entonces a la resolución de problemas, utilizando -por comodidad- dos casos de ver de "La infancia de un jefe".

Problema 1: Lucien, que es el personaje central de la obra, cree de niño que hay una conspiración contra sus padres. Un hechizo habría convertido a su padre en su madre,

y a su madre en su padre (pp. 7-8). En nuestro contexto, esto no es otra cosa que un caso de alguien que viendo a su madre, esa persona que ve, es su madre o no lo es.

Glosario	Esquematización	Solución
p: Lucien	Vp q=P ├ q=P v q= -P	-1 Vp q=P
q: Sra. Fleurier		Va b=P RS1-1 1 (p/a, q/b)
P: ser madre de Lucien		3 b=P RR2 2
		4 b=P v b= -P Iv 3
		5 q=P v q= -P RS2 4 (b/q)

Y ¿qué se ha obtenido?. Dicho en pocas palabras, demostrar que si alguien dice lo que figura en 1, tiene derecho a afirmar lo que figura en 5.

<u>Problema 2</u>: En la p. 11 de “La infancia de un jefe”, Sartre le hace vivir a Lucien su primer gran frustración. Convertido ya en un adolescente, deja de ser el objeto de las atenciones de los mayores; ahora parecen no verlo. De ahí, él deduce que para ellos ha dejado de existir.

Se puede decir entonces que estamos frente a un caso del tipo: alguien no ve a una determinada persona, luego, esa persona no existe.

Glosario	Esquematización	Solución
p: Sr. Jules	-Vp q ├ q= -P	-1 Vp q
q: Lucien	2 Va b RS1 1 (p/a, q/b)	
P: ser existente	3 Va b=P RE=1-1 2	
	4 Va b= -P Def. 3	
	5 b= -P RR2 4	
	6 q= -P RS2 5 (b/q)	

¿Lo obtenido?. Lo mismo que en el caso anterior. Aceptado lo que figura en 1, se tiene derecho a sostener lo que figura en 6.

4 Conexiones

Hasta aquí, la lógica. En lo que sigue, me ocuparé de otra cosa, me ocuparé de demostrar que esto de la lógica del ver no se queda en un mero divertimento para lógicos. Dicho con otras palabras, intentaré responder a la pregunta ¿para qué sirve esta lógica?.

En verdad se me ocurren algunas cosas, pero aquí, lo que puede interesar es la conexión con la psicología. Apuntemos lo siguiente. Primero que nada, a la psicología no puede importarle en demasía el procedimiento formal de esta lógica del ver. Lo único que le es posible pedirle a esta lógica es que en verdad demuestre lo que dice demostrar y que eso le sirva para algo. El mejor de los casos, transformarla en una ciencia auxiliar de la psicología. Pero como es evidente, eso no significará 'tomar' a toda la lógica, sino lo que para la psicología es relevante. Y ¿qué cosa de esa lógica puede ser relevante?. Sin duda, las inferencias ya demostradas por ella y sus reglas. Expliquemos esto.

En relación a las inferencias. Lo que es viable a este respecto es efectuar correlaciones entre inferencias demostradas y determinadas psicopatologías. Por ejemplo, correlacionar el problema 1 con un delirio persecutorio.

Ahora las reglas. Las reglas no se limitarían a ser normas que permiten deducir, son mucho más. Tienen que ver con ciertas operaciones fácticas que un individuo realiza. Para explicarlo, sigamos con el ejemplo que veníamos manejando. Allí, según vimos, se aplica RR2. Esto no es un asunto menor; al contrario, es la expresión de un modo de ser que no se percata ni de la problematicidad de pasar de ver que una cosa es de tal manera a sostener que efectivamente es de tal manera, ni quizá, de la diferencia entre una y otra afirmación.

Aceptado esto, falta aún determinar qué cosa haría la psicología con esa información. Y, lamentablemente, aquí la respuesta es únicamente hipotética. Mi hipótesis es la siguiente:

- es posible 'abrir' una línea de investigación, en la que la psicología -o al menos una parte de ella- sea concebida como una ciencia que tiene por objeto los modos de construcción del ver;
- hacer trabajar a esa psicología con un nuevo enfoque terapéutico, basado en una reeducación del ver. Ilustremos este punto.

Imaginemos la siguiente situación. En la sala de espera del consultorio de un terapeuta entra un futuro paciente (para seguir con el ejemplo, digamos que se trata de Lucien) y la recepcionista le entrega una cierta cantidad de hojas para llenar allí mismo. En esas hojas se pide el nombre, la dirección, el teléfono, etc., pero además habría otras cosas: cuestionarios y/o test. Imaginemos ahora que a Lucien se le presenta esto: Piense en una persona que le haya hecho unas cuantas maldades. Ahora conteste honestamente con un 'si', un 'no', o un 'tal vez' a esta pregunta: ¿es mala esa persona?.

¿Qué respuesta podemos esperar de Lucien?. Si es honesto, la respuesta ha de ser un 'si'. ¿Qué es lo que se obtuvo?. Un diagnóstico del modo de ver de Lucien. Y ¿de qué le serviría eso al terapeuta?. Simplificando mucho las cosas, le serviría para realizar un prediagnóstico de Lucien. El terapeuta sabría que quien padece delirio persecutorio tiene ese modo de ver; luego vendría el trabajo dentro del consultorio, trabajo

con el cual se confirmaría o no ese prediagnóstico inicial. La ventaja de esto es evidente; por un lado, imprime una dirección al diagnóstico, y por otro lado, puede acortar los tiempos.

Pero esto puede pensarse mucho más allá del diagnóstico, es posible pensarlo en relación a la terapia misma. La cuestión sería reeducar el modo de ver de Lucien. Si se le enseña a ver de otra manera, tal vez encontremos la solución para su delirio. Esto, sin embargo, es una fuerte sospecha de la cual no creo que sea conveniente abrir opinión por el momento. Creo más bien que esto tiene que discutirse alrededor de una mesa en la que estarían sentados, como mínimo, lógicos, filósofos y psicólogos. Dicho esto, Señores, tienen la palabra.

Notas

(1) "Conciencia moral y acción comunicativa", La filosofía como vigilante e intérprete, p. 28.
(2) "El tema de nuestro tiempo", pp. 100-101.
(3) Ávila, en "La observación, una palabra para desbaratar y re-significar", sostiene esto mismo, pero lo hace depender de la cultura. Para él, la preeminencia de lo visual es producto más de la denominada cultura occidental.
(4) "La evolución de mi pensamiento filosófico" p. 149.
(5) "Saber y creer", p. 63.
(6) Por cierto, esto es algo que un gestaltista podría discutir. Podría argumentar que **a** ve a **b** como un 'todo' y no de 'a partes'. No obstante y aún cuando pueda tener razón en esto, de ningún modo puede llegar a decir que las partes sean in-visibles.
(7) "Lógica simbólica", p. 158.
(8) Esta regla puede fundamentarse de la misma manera que las reglas de introducción y de eliminación de los cuantificadores en la lógica de predicados: si la expresión vale para un individuo sin especificar (o especificado), también ha de valer para un individuo específico (o sin especificar).

Referencias

Ávila, Rafael (2004) "La observación, una palabra para desbaratar y re-significar". Cinta de Moebio N° 21, http://www.moebio.uchile.c1/21/frames01.htm. Facultad de Ciencias Sociales. Universidad de Chile. Chile

Garrido, Manuel (1974) "Lógica simbólica" Edit. Tecnos, Madrid

Hahn, John F. (1979) "Introducción a la psicología", Edit. Psique

Hintikka, Jaakko (1979) "Saber y creer" Edit. Tecnos, Madrid

Nagel, Ernest (1974) "La lógica sin metafísica" Edit. Tecnos, Madrid

Ortega y Gasset, José (1976) "El tema de nuestro tiempo" Edit. Revista de Occidente, Madrid

Rovaletti, Lucrecia (2006) "Esquizofrenia, sentido e insensatez", "Relaciones" 263, Montevideo

Russell, Bertrand (1976) "La evolución de mi pensamiento filosófico" Alianza Editorial, Madrid

Sartre, Jean Paul (1994) "La infancia de un jefe", Alianza Editorial, Madrid

Ontologías y agentes de red: Un recambio para la I.A. clásica

Enrique Alonso y Javier Taravilla

Dept. de Lingüística, Lenguas Modernas, Lógica y Filosofía del Ciencia
U.A.M.
enrique.alonso@uam.es, javier.taravilla@uam.es

Resumen. Este trabajo puede ser fácilmente considerado como una de las muchas secuelas provocadas por el artículo seminal de Berners-Lee et al. en el que se establece la hoja de ruta de la nueva Web, o lo que es lo mismo, la Web semántica. Aunque no tenemos ningún inconveniente en ubicar nuestra contibución dentro del ámbito de ese proyecto, lo cierto es que nuestros objetivos no son los que cabría esperar de un seguidor fiel de la doctrina fijada por el W3C Consortium. Esto se debe a que el recambio del trabajo tradicional sobre Inteligencia artificial, ha venido representado en el marco tecnológico por la aparición de los sistemas expertos, sistemas agentes y finalmente sistemas multiagente o agentes de red. El objetivo de este árticulo es hacer una propuesta de sistema agente que ayude al tratamiento y recuperación de documentación en instituciones que generen grandes cantidades de documentos, y que su accesibilidad y consulta sea esencial para la buena marcha de la organización. LOCUS es una propuesta de sistema agente "cerrado" que se presenta como proyecto piloto para resolver y abordar con los avances informáticos, problemas de este calado.

Palabras clave: Sistema Experto, Sistemas Multi-Agente, Interacción y Consenso, Ontologías, Web semántica y LOCUS

1 El estado de la cuestión

El recambio para los proyectos de Inteligencia Artificial *Clásica* ha venido representado en los últimos 20 años por los *Sistemas Expertos* [1] y como evolución a estos, los llamados *Sistemas Multi-Agente* [2] conocidos también como MAS. En este último grupo el tema de qué sea una Ontología de Software y las comunicaciones entre ellas se presenta como nuevo campo para analizar.

1.a. Un enfoque distinto: los sistemas expertos

Como *Sistema Experto* entendemos un programa de ordenador que realiza la tarea de un experto humano. De aquí su obvia caracterización como *experto*: no se busca sino ahorrar con él la necesidad de tener que recurrir a los costosos servicios que ofrece el especialista humano de un campo, cada vez que necesitamos decidir sobre alguna

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 65-78, 2006.

cuestión de esa materia. Podriamos definirlos de modo muy general, como un sistema informático que, basado en ciertas reglas de inferencia y programación, diera consejos de modo muy similar a como haría un experto humano: ayuda para realizar diagnósticos, cálculos o tomar decisiones. En este área fue conocido en las década de los 80 y los 90, el programa MYCIN usado para detectar enfermades de tipo bacteriano en sangre [3]. Ya con anterioridad y bajo esta perspectiva, una calculadora venía a ser un "sistema experto" en lo que a realizar cálculos se refiere. Nuestra propuesta es un buscador documental que almacenando los archivos de un modo determinado, permita recuperar datos la información guardada de forma "natural" y eficaz. Entendemos que sigue el espíritu de aquellas propuestas de desarrollo "experto", teniendo en cuenta los últimos trabajos en lo que a sus sucesores se refiere.

Por otro lado y para intereses de IA clásica, los programas de este tipo vendrían a franquear los límites y alcance de lo que entendemos como "juego de la imitación" a la vez que abren de nuevo el debate sobre lo que la simulación puede suponer acerca del carácter inteligente o no de nuestros ordenadores. Este debate, de indudable interés, no es el que nos urge ahora tratar.

1.b. Agentes y sistemas multi-agente (MAS)

El testigo a estos sistemas vino con los llamados *agentes* y *sistemas agentes*. Un *agente* no es sino una entidad que se desenvuelve con cierto grado de autonomía en un entorno determinado, siendo un término que pasará a usarse en la jerga informática. Como definición general de *agente* encontramos, Wooldridge y Jennings [4], Wooldridge [5], y su aplicación al software en Symeonidis y Mitkas [2]. Surgen entonces definiciones que catalogarán de *agentes* a las computadoras, y ciertas *entidades* de software podrán ser llamados *agentes de software (software agents)*: son aquellos programas capaces de alcanzar una serie de objetivos o arrojar un juego de resultados en sus relaciones con otras entidades, sean humanas o mecánicas. Esta serie de definiciones no aclaran mucho la situación, al poder reconocer bajo la categoría de *agentes de software* a una gran cantidad de aplicaciones informáticas. Se empieza en los últimos años por tanto a hablar de *Sistemas Multi-Agente* o MAS (*Multi-Agent System*) como evolución de aquellos sistemas expertos anteriores, que se diferencian de estos en su grado de complejidad.

Un *Sistema Multi-Agente* se caracteriza por querer probar que problemas que se entienden competencia de agentes disribuidos y requieren de la sinergia de un cierto número de elementos dispares, pueden ser resueltos eficientemente por estos *Sistemas*. La complejidad de un MAS dependerá del número de agentes que se vean implicados. Las arquitecturas MAS superan las propuestas de agentes que se insertan en entornos distribuidos, al ser coordinados por grupos de trabajo en contacto (*Agent Working Group*), a la vez que están en constante comunicación e intervención con otros agentes o tipos de agente. Vemos por tanto que la colaboración y relación entre hombres y máquinas es punto fundamental en los últimos desarrollos de Inteligencia Artificial. La frontera entre máquinas y hombres se desdibuja y el trabajo cooperativo entre entidades se muestra esencial.

Se habla a su vez de agentes singulares cuando solo hay implicado uno, a la vez

que encontramos MAS de mayor elaboración cuando hay un mayor número de agentes cooperando. Toda esta evolución surge de las arquitecturas distribuidas de Software, la computación en red, la iniciativa de Servicios Web y la llamada Computación Distribuida, temas todos ellos trabajados intensamente en los últimos cinco años.

1.c. Ontologías para el software

Y es aquí donde queremos ir a parar: una vez presentados los sistemas multi-agente, empieza a surgir el problema de comunicación entre ellos; cada programa o sistema es hecho conforme la experiencia como programador de su creador. Dicho de otro modo: para que los MAS informáticos funcionen es básico que la comunicación establecida entre esas entidades de software sea lo más correcta posible. Recordamos que para el desarrollo de un agente múltiple se necesitaba la intervención de varios agentes de forma coordinada. Aparece entonces el debate entorno a las *Ontologías de Software,* que no son sino el modelo tencológico y computacional de lo que tradicionalmente se ha entendido como ontologías.

Una *Ontología* puede ser definida como una especificación de los conceptos que un software agente usa. Es una descripción de los conceptos y relaciones que pueden existir en un agente informático [6]. Para aquellos más cercanos a la lógica podemos decir que no es sino una pequeña descripción del mundo en Lógica de Primer Orden que divide los conceptos a utilizar en distintas clases, con sus respectivas características y establece un sistema de relación entre ellos. La relaciones entre conceptos son las de *subclase, superclase, equivalencia ≡, diferencia ≠,* o *similitud ≈.* A la vez la OWL establece 11 axiomas de relación entre los conceptos y objetos de una Ontología [7]. Ofrecen al agente una interpretación para los mensajes que recibe y un uso que debe darle a las palabras. Es aquí cuando sirve diferenciar entre Sistemas Multi-Agente Abiertos y Sistemas Multi-Agente Cerrados. El proyecto LOCUS representaría un ejemplo de Sistema Cerrado, por ahora. La diferencia entre ambos es que los cerrados son aquellos que trabajan con los significados de una única ontología encontrando todos sus elementos claramente definidos, mientras que los abiertos son aquellos donde los agentes tienen que comunicarse con varias comunidades que usan diferentes ontologías[8]. Un ejemplo de estos sería un buscador de biblioteca o un procesdor de textos cuando trabaja con un tipo de archivos conocidos para él, como ejemplo de cerrado; cuando trabajamos con archivos *.Pdf* y archivos *.doc* o su versión abierta *.rtf*, podemos hablar de de distintas ontologías abiertas a la hora de trabajar con los formatos de texto, por poner ejemplos de uso común. El problema de las compatibilidades semánticas está en aquellos sistemas que no comparten la misma ontología.Y es aquí donden surgen los estudios de traducciones y movimientos entre Ontologías. En esta línea Java parece ser el lenguaje puente con más posibilidades[9].

1.d. Situación de trabajo

La principal diferencia es que nuestro problema no es la recuperación de información y la interacción de agentes en el entorno fijado por la Web sino en dominios corporativos que, por lo general, trabajan con la vista puesta en otros fines. No nos preocupa

especialmente si los clientes potenciales de una compañía son capaces de llegar a su página a partir de criterios de búsqueda más o menos pertinentes, o si podemos contar con un agente de red que realice ciertas tareas basadas en la información que en ellas se encuentra. Todos estos problemas tienen un ámbito mucho más amplio que el que nos interesa aquí.

Las empresas e instituciones para las que trabaja el común de los mortales producen para su correcto funcionamiento una gran cantidad de documentación generalmente contenida en archivos de texto. Los formatos de texto –nos cuidamos de llamarlos formatos .doc, por entender que .doc es la nomenclatura propietaria de dichos entornos, pudiendo reconocer los archivos de texto bajo otras tantas denominaciones [10]– representan el modo que entidades que generan un gran número de documentación tienen de almacenar sus decisiones, fijar sus directrices y comunicar información relevante a los distintos puntos de su estructura. Es cierto que en el intercambio de información el correo electrónico ha tomado la delantera al archivo tradicional de texto, pero sólo hasta cierto punto y con las severas limitaciones que todos conocemos.

Mientras que la Red ha creado una considerable cultura de la búsqueda de información, no vemos nada parecido en el entorno de la producción de texto. Mucho menos en el estudio de formas estructuradas de almacenamiento de textos, de modo que su recuperación posterior sea mucho más agradable y esté informátizada. Se puede alegar que esta especie de desfase es sólo circunstancial y que encontrará adecuada solución y respuesta en el momento en que corporaciones de todo tipo adopten para su uso un formato HTML –XML, en realidad–. Tenemos dudas de que esto sea lo más indicado o que represente siquiera el curso posible de los acontecimientos. Los textos son, desde el punto de vista de la información que contienen, entidades mucho más densas y complejas que el código en HTML. La relación que el usuario establece con un documento HTML no parece la misma que la que establece con uno de texto llano. Sería muy arriesgado por nuestra parte entrar ahora en detalles acerca de unas diferencias que cambian constantemente, pero parece claro al menos que el objetivo de un buscador de red no es el mismo de un buscador de tipo documental. En el primer caso se intenta llegar a aquellas páginas que contienen los términos claves de la búsqueda en determinadas ubicaciones, pensamos, por ejemplo, en las palabras claves que figuran en ciertas *tags* o etiquetas en las cabeceras de los documentos HTML. Y no parece que esta doctrina se pueda aplicar a la documentación corporativa sin más.

Sorprende que mientras que el códigos HTML ha tenido desde sus propios inicios la vista puesta en la accesibilidad de la información y en facilitar al máximo la forma que el usuario accede a la misma, el dominio de los documentos de texto se haya mantenido tan fiel a criterios francamente obsoletos. El uso de los recursos informáticos en la producción y edición de texto permanece, aún hoy, fuertemente ligado al intento de sustituir con éxito a la máquina de escribir o, por ser generosos, a la pequeña imprenta. Y lo ha conseguido, pero al precio de hacer que la diferencia entre un papel impreso y el documento sobre la pantalla sea realmente poca por lo que hace a la localización de información relevante.

Responder a preguntas acerca de cuándo se adoptó esta o aquella decisión, qué se dijo en relación a tal o cuál cuestión, quién se quedó encargado de la ejecución de un cierto proyecto o qué fecha fue marcada como límite para entregar una determinada

documentación, sigue siendo hoy tan difícil como lo era antes de la introducción generalizada del ordenador en las empresas e instituciones. Nuestro objetivo es analizar si las intuiciones que acompañan los estudios en torno a *Agentes de software* y las *Ontologías* usadas pueden ser útiles para mejorar en algo esta situación.

2 La Filosofía de los lenguajes de marcas

La sugerencia más valiosa de las muchas que acompañan al proyecto de la Web semántica es la intensa reflexión promovida en torno a lo que podríamos llamar texto etiquetado o anotado. El trabajo realizado en los últimos tiempos en torno a los lenguajes de marcas *–mark-up languages-* ha venido a mostrar la gran utilidad que para la gestión automática de la información puede tener el uso de metadatos elaborados de diversos lenguajes de marcas. La proliferación de sistemas distintos ha forzado al Consorcio de la W3C a proponer un estándar, XML, al que deberían ajustarse todas las iniciativas en esta dirección. El tiempo dirá si se logra.

Pese a lo que pueda parecer, el uso de etiquetas concebidas como metadatos que acompañan a un documento sin formar parte de su presentación final no es cosa de ahora. De hecho, es el sistema que fue originalmente empleado por los procesadores de texto más comunes para incorporar formato al texto plano. Todos sabemos que desde el triunfo de las llamadas GUI [11] se ha perdido la distinción entre texto y formato al ofrecerse presentaciones finalistas en pantalla de nuestros documentos. Y quizá debamos empezar a ver en esto una cierta pérdida más que una definitiva ventaja.

La diferencia entre el uso primitivo de las etiquetas que empiezan a ser incorporadas como metadatos, y el que ahora se discute es sustancial. En los procesadores de texto a que hemos hecho mención las etiquetas forman parte de la herramienta no pudiendo modificarse de ningún modo. No son un elemento que pueda ser importado desde el exterior porque no hay un lugar desde el que hacerlo. No existen los lenguajes de marcas como entidades independientes. Parece que pese a haber existido casi desde siempre, es solamente ahora cuando hemos empezado a considerar la importancia de unos objetos, los lenguajes de marcas, que ocupan un nicho muy particular entre las especies del software contemporáneo. No cabe duda de que en el futuro habremos de estudiar mucho más a fondo sus propiedades de las cuales, por cierto, sabemos poco ahora.

El uso que los lenguajes de marcas tienen en la actualidad tiene que ver con lo que podríamos considerar como el aumento del valor de un documento estándar. Se trata, precisamente, de añadir valor a un documento al margen de su contenido explícito o final. Esta estrategia se conecta con un tema que nos interesa por razones quizá independientes: la vigencia del proyecto de la IA. Se supone que la IA tiene entre sus objetivos el diseño de herramientas capaces de interpretar el sentido de los textos que los seres humanos producimos con el fin de desarrollar una conducta coherente con ese contenido. Pero el caso es que la IA llega tarde a su cita. Esta es la brecha por la que ha penetrado con fuerza el proyecto de las Ontologías y la Web semántica. Si no somos capaces de hacer que un agente artificial interactúe significativamente con nuestros documentos, habrá que resignarse a incorporar en ellos instrucciones que les

indiquen qué hacer con ellos. Esta información que acompaña al contenido textual de un documento y que, propiamente no está destinada a formar parte de su contenido explícito –el que, por ejemplo, puede ser impreso en papel– es la que queda para el uso de metadatos elaborados con la ayuda de lenguajes de marcas. Aunque puede no parecer algo demasiado radical en principio, lo cierto es que supone un cambio sustancial en alguna de nuestras actitudes frente al documento en formato electrónico. Hasta ahora la elaboración de un texto en formato electrónico no difería en nada de la que tendríamos, caso de disponer tan solo de una antigua máquina de escribir. El texto compuesto es el que finalmente ven sus potenciales destinatarios. El uso de metadatos obliga –aconseja en todo caso– a actuar de otro modo. La elaboración de un texto tendría dos momentos seguramente solapados. Uno, en el que se compone en texto explícito o final, y otro en el que se incorpora sobre el mismo, en forma de metadatos, información relevante destinada a indicar a otros usuarios cómo recuperar de forma eficiente lo que allí se dice o qué hacer con los datos que en él se aportan. Aquí es donde nace una suerte de *simetría.* Es el usuario, quien como productor de documentos, debe incluir en ellos información acerca de cómo debe almacenarse su contenido. Con esto no se busca sino que en el momento de recuperar una información, la responsabilidad de su hallazgo se deba tanto al preguntar eficaz del agente humano, como a la ayuda y comodidad ofrecida por el agente mecánico. Esta facilidad útima viene establecida por sistemas de almacenamiento previos, que hemos *consensuado* con toda persona que archive o guarde un documento.

Esta forma de trabajar supone retos innegables ya que su éxito depende en buena medida de nuestra habilidad para diseñar métodos de trabajo que hagan fácil este doble proceso de incorporación de información. No tiene mucho sentido pensar en dos procesos, uno de elaboración tradicional del texto, y otro de incorporación de metadatos, ya que si ha de ser así, simplemente no será. Este cambio de actitud frente a nuestros productos, y la responsabilidad que adquirimos al adjuntar los metadatos que los acompañan es algo de lo que seguramente volveremos a hablar en breve.

3 Ontologías de campo y ontologías idiomáticas: El Proyecto LOCUS

Los lenguajes de marcas empleados para incorporar metadatos en nuestros documentos reciben, como es bien sabido, el nombre genérico de *ontologías documentales*. Por nuestra formación filosófica bien podríamos iniciar ahora una docta exposición acerca del uso previo de este término y de su relevancia dentro de la filosofía tradicional, pero lo cierto es que no merece mucho la pena detenerse en esto.

El uso de *ontologías de tratamiento de texto* parece localizado, al menos por ahora, en la producción de documentos HTM, XML, y RDF. Se trata, en definitiva, de objetos mayoritariamente destinados al consumo de la Red. El lenguaje de marcas que parece haber ganado terreno en estos años es OWL[7] en lo referente a ontologías de tipo general y RDF dentro de los medios de comunicación. Ambos aportan metaestructuras en las que albergar los mapas conceptuales que permiten interpretar las relaciones que guardan entre sí los términos que figuran en una página web. Si con-

sultamos la documentación que aporta la página del Consorcio de la W3C que mencionamos anteriormente vemos que el desarrollo de ontologías es especialmente intenso en ciertos dominios médicos –oncología, más concretamente- y académicos. Se trata en ambos casos de un uso temático del texto anotado destinado a comunidades que puede compartir estándares técnicos ampliamente aceptados. Los problemas que se plantean con esta maniobra se pueden apreciar, no obstante, ya en este punto. La incorporación de una ontología en una página web, portal o documento, tanto da, no es un acto del todo inocente. Una ontología destinada a fijar el mapa conceptual de un cierto ámbito exige un *consenso* que rara vez se da en las comunidades humanas. Por otra parte, una *ontología temática* sólo fija el uso de un término relativamente a un determinado idioma dejando la traducción más conveniente a otros equipos más o menos coordinados. Puesto que el problema del idioma fija el límite del intercambio eficiente de información, nos concentraremos en el primer aspecto del problema.

Una ontología es, no se olvide, un mapa conceptual que intenta explicar las relaciones entre los términos que aparecen en un cierto entorno, lo que podría denominarse el *alcance* de esa ontología. ¿Qué sucede cuando páginas afines no comparten la misma ontología? ¿Es posible que dos ontologías entren en conflicto creando confusión?[12] Hasta ahora el desarrollo de la Web se había basado exclusivamente en la suma directa de todas las aportaciones de sus usuarios. No había, por así decir, ningún criterio que exigiera una coherencia al conjunto. No podía haber conflicto entre dos páginas distintas. El desarrollo de la Web semántica rompe si no en la letra, sí al menos en su espíritu, con este modo de operar tan valorado por todos nosotros. El uso de ontologías impone como frontera natural el de la comunidad que la comparte y dista mucho de estar claro cuál es el mejor modo de gestionar este tipo de arancel nuevo en el desarrollo de la Web.

Tampoco es de recibo la pugna que evidentemente se puede establecer por la imposición de una ontología en un dominio, ni la lengua en que esta se encuentre elaborada.

Todas estas situaciones afectan básicamente a las ontologías temáticas concebidas como medio para facilitar la navegación semántica en una Red concebida hasta ahora como un medio puramente sintáctico. ¿Caben otras opciones?

Los documentos de texto contienen información sobre la que habitualmente nos preguntamos. Queremos saber, por ejemplo, quién hizo algo, o cómo lo hizo, qué se decidió o cuándo hay que realizar un determinado trámite. En los últimos años han ido en aumento las investigaciones y desarrollos de todo aquellos que podemas hacer con los ordenadores, fijando especial atención a los temas de interacción con las máquinas. Como conseguir, a través de aplicaciones y sobre todo de interfaces, que distintos usuarios pueden tener un uso cada vez más eficaz con estas entidades tecnológicas: aplicaciones a docencia y comunicación, agrimensura, aparición de robots, usuarios domésticos o grupos con discapacidades son sectores en los que se ha trabajado el grado de relación entre humanos y máquinas [13]. El Proyecto LOCUS va dirigido a usos de este tipo en instituciones que precisen un tratamiento eficaz de su documentación, teniendo en cuenta que muchos organismos e instituciones, en especial las académicas o gubernamentales, se caracterizan en gran medida por la documentación que producen.

Para obtener información sobre todos esos extremos, empleamos en castellano el tipo de recursos que la lengua posee, es decir, los adverbios y locuciones que nos permiten formular preguntas. Nuestra propuesta es dirigirnos a estos recursos antes que a los propiamente semánticos con el fin de establecer una estructura de metadatos eficiente desde el punto de vista de recuperación de la información. Hace años se habló mucho de una *lógica erotética* para analizar la estructura de las preguntas que pueden hacerse en una lengua, no importa cuál. Nuestra propuesta es elaborar un lenguaje de marcas de tipo erotético que permita al usuario marcar la información de sus textos de forma que otros puedan recuperarla del modo en que el primero desea que se haga. Nuestra propuesta recibe el nombre de LOCUS y tiene la estructura que se describe a continuación.

A. Elementos del sistema (elementos simples)

<quien (prep=”...”)>...</quien>
<que (prep=”...”) (rasgo={s,v})>...</que>
<cuando (prep=”...”)>...</cuando>
<donde (prep=”...”) >...</donde>
<como (prep=”...”) >...</como>
<cual (prep=”...”)>...</cual>
<cuanto (prep=”...”)>...</cuanto>

Todos estos elementos tienen funciones gramaticales bien definidas. Aquí representan, no obstante, las respuestas a preguntas encabezadas por las formas gramaticales correspondientes.

Admiten como modificadores distintas preposiciones. Esto permite alterar sus funciones dentro de la oración. El modificador prep=”...” indica la posibilidad de incluir una preposición que modifique la pregunta a la que es respuesta el texto marcado.

> “El Decano nombró director de la sección X al Sr. Y” quedaría como sigue:
>
> <quien> El Decano </quien>
> <vrb> nombró </vrb>
> <que> director de la sección X</que>
> <quien prep=”a”> al Sr. Y </quien>

donde el elemento

<vrb>...</vrb>

corresponde al núcleo oracional y no admite modificadores.

Es importante insistir en que estos elementos no conservan necesariamente su función gramatical, pero no merece la pena entretenerse ahora en ello.

B. Elementos del sistema (elementos complejos)

```
<oracion id=”n”>
        <vrb>...</vrb>
</oracion>
```

Dentro de este elemento se pueden y deben incluir los anteriores del modo en que se indica en el siguiente ejemplo:

“La Administradora decidió que las secretarias de Departamento realizaran la matrícula los días 7 a 8 de abril”

```
<oracion id=”1”>
        <quien>La Administradora</quien>
        <vrb>decidió</vrb>
        <que >que
                <oracion id=”2” >
                        <quien>las secretarias <que prep=”de”>de
                                        Departamento</que></quien>
                        <vrb>realizaran</vrb>
                        <que >la matrícula</que>
                        <cuando>los días 7 a 8 de abril</cuando>
                </oracion>
        </que>
</oracion>
```

Esta anotación es algo exagerada. Nada obliga a ser tan exhaustivo. Piénsese que siempre prima el deseo del usuario.

El modificador id=”n”, donde *n* es un entero positivo, es obligatorio en este elemento, ya que de otro modo se producirían ambigüedades.

El orden en que se muestran los elementos simples que forman el elemento compuesto *oración* se puede alterar sin problemas. Se trata de preservar el orden habitual en que se redacta una oración y es a eso a lo que debe ajustarse el mecanismo de etiquetado de texto.

C. Localizadores

```
<documento id=”n”>...</documento>
<seccion id=”n” nombre=”...”>...</seccion>
```

El primer elemento se aplica a todo el texto, por tanto sólo puede haber uno por documento. Da inicio al documento y figura al final del mismo. El elemento *sección* puede repetirse dentro del texto, y puede anidarse.

D. Deícticos.

```
<item id=”n”>....</item>
</item tipo=”n” id=”m”>
```

La primera forma asigna contenido al elemento *item*. La segunda es un elemento vacío que se emplea para señalar una ocurrencia del item *n*. Se emplea el modificador *tipo* para aludir al fragmento de texto representado y el modificador *id* para identificar la ocurrencia de esa referencia al item *n*.

El Decano presentó <item id=”n”>el informe</item> a la Junta

<quien>Su contenido <que prep=”de”></item tipo n id=”m”></que></quien>... alarmó a los presentes.

Contar con un medio para anotar texto previamente marcado es especialmente importante en la primera operación, es decir, en la que se identifica el contenido del item, mientras que para incluir menciones a ese item es imprescindible poder visualizar rápidamente todos los que están disponibles en el texto.

Esta opción presenta un inconveniente estético. El proceso de identificación de un item puede contravenir el principio general de que los elementos raíz, es decir, los que pueden figurar de forma autónoma sean solo elementos *oración*. Al marcar como item una oración contradecimos esa norma. La otra opción es establecer un mecanismo por el cual se pueda mencionar cualquier fragmento marcado haciendo uso de elementos vacíos. Esto es fácil en el caso de oraciones, que incluyen de forma necesaria un identificador, pero no lo es tanto en el caso de elementos simples que heredan los de la oración de la que forman parte. “Eso resultó ser falso” puede quedar como sigue:

```
<oracion id=”n”>
        <que >Eso </oracion id=”m”></que>
        <vrb>resultó ser</vrb>
        <que>falso</que>
</oracion>
```

donde </oracion id=”m”> apunta a la oracion con identificador *m*.

Por el momento se puede mantener la ambigüedad hasta comprobar qué uso es el más adecuado.

A todos estos elementos aún se pueden sumar otros destinados a conectar entre sí oraciones. En este apartado entran los consabidos operadores lógicos y otros destinados a indicar funciones no estrictamente lógicas, como por ejemplo, la explicación informal. Omitiremos aquí su descripción.

Si comparamos esta propuesta con otras ontologías apreciamos, en primer lugar, que LOCUS no es propiamente una ontología, sino un lenguaje de marcas de tipo muy general. No se establece un modelo del mundo ni un mapa categorial del tipo que se obtiene, por ejemplo, empleando OWL. Se trataría, si se nos permite, de una

ontología genérica ligada a la lengua y por tanto no se vería sometida al tipo de obstáculos que hemos comentado líneas atrás. Pero nótese que nuestra propuesta sí supone aceptar una cierta *ontología* o estructura a la hora de archivar documentos, basándonos en las funciones de los elementos gramaticales del texto. Aceptamos almacenar la información de un modo determinado, asumimos una estructura muy general de los textos, por lo que a parte de *genérica*, trabajar con esta herramienta supone aceptar una cierta ontología *particular*. En un documento anotado en LOCUS el usuario no aporta información acerca del contenido semántico de su documento sino una guía acerca de cómo recuperar la información que acaba de redactar.

4 Preguntar sólo lo que tiene respuesta

Para entender este punto conviene reflexionar un poco acerca del modo en que es posible recuperar información de un documento o conjunto de ellos. Supongo que lo que nos gustaría es poder hacer una pregunta muy general como las que normalmente se producen en intercambios típicamente humanos. Estamos aún lejos de esto. La otra opción es la búsqueda sintáctica basada en palabras clave. Ya sabemos los límites que tiene esta estrategia. La tercera opción es intentar preguntar sólo aquello que en principio tiene respuesta, intentando, además, que ésta sea clara y concisa. LOCUS permite plantearse de forma efectiva este objetivo. Es por alumbrarlo con una metáfora, una estrategia tipo interrogatorio *policial* o de juego de *aventura gráfica* que mencionaremos. Al igual que cuando un agente del orden o abogado en un juicio quieren extraer información de un posible delincuente, saben qué han de preguntar y cómo. Nosotros hemos de hacer algo muy parecido con nuestro buscador. En base a la estructura con la que hemos introducido la información, ésta está almacenada, por lo que el saber qué preguntar será esencial para el éxito de nuestra tarea.

Un texto anotado con LOCUS ofrece la opción de almacenar los contenidos de las etiquetas incluidas dentro del elemento <oracion id= >...</oracion> en contenedores apropiados. Todos los verbos que aparecen en las oraciones de un documento y que están dentro del elemento <vrb>...</vrb> serán reunidos dentro de la categoría verbos. Las expresiones que han sido etiquetadas mediante el elemento <que prep= >...</que> son alojados en el correspondiente contenedor, repitiéndose la operación para cada una de las etiquetas disponibles. Esa misma operación tendría lugar en cada uno de los documentos que forman el alcance de nuestro buscador formándose de este modo un entorno de búsqueda relativamente acotado.

La elaboración de una pregunta tiene ahora el aspecto de una típica selección de ítems entre los distintos contenedores que suministra el buscador, los cuales han sido previamente cargados con todos los documentos que caen bajo su alcance. Esta técnica recuerda en algo al método empleado en aquellos videojuegos que prevén diálogos interactivos entre los personajes o entre el jugador y estos últimos. A partir de un panel de órdenes básicas se despliegan posibles acciones generalmente previstas por el juego. Las pruebas que se han realizado hasta el momento manejando este entorno para la elaboración de preguntas resultan bastante esperanzadoras. La visualización de los ítems alojados en los distintos contenedores, permite elaborar preguntas bastante orientadas a posibles respuestas. Los casos de preguntas fallidas son por ese

motivo relativamente bajos. Al visualizar las opciones existentes en cada categoría el usuario busca de forma natural aquellas que mejor se adaptan a su pregunta primitiva. La única dificultad que se puede presentar es una excesiva proliferación de ítems dentro de cada categoría o contenedor. Y aunque lo cierto es que no disponemos de datos que permitan valorar el posible impacto de esa circunstancia, no es de prever que esta circunstancia suponga de inmediato un riesgo sustancial, ya que el vocabulario que suele emplearse en contextos profesionales resulta sorprendentemente reducido por lo general. Como parece evidente, hay mucho más que decir a este respecto de lo que estamos revelando en estas líneas, pero estamos seguros que se puede suplir con un poco de imaginación.

Las respuestas que puede obtener una pregunta elaborada por este procedimiento pueden ser tan concretas o amplias como se quiera. Si fue el Decano quien decidió que se nombrase a fulano como responsable de tal o cual sección y la pregunta fabricada ha sido "¿Quién fue quien decidió que se nombrase a fulano responsable de aquella sección?" es obvio que la respuesta será "El Decano". Pero podemos abrir el foco de nuestro objetivo admitiendo más de una opción en cada categoría. Podemos no saber si el Decano tomó la decisión o simplemente sugirió que se tomara. Entre los verbos disponibles pueden figurar ambos y nada impide que hagamos una pregunta compleja del tipo ¿Quién sugirió o decidió que...?

Tampoco debemos evitar que se hagan preguntas absolutamente generales del tipo "dime todo lo que se sepa de tal...". En este caso la similitud con las búsquedas típicamente sintácticas será evidente, dependiendo su eficiencia del volumen de información disponible. Nuestra experiencia es que los usuarios suelen elaborar preguntas concretas si disponen de suficientes ítems en cada una de las categorías que suministra LOCUS.

5 Agentes de red y lenguajes de marcas: La definición de tareas

La parte menos elaborada de nuestro proyecto es la que tiene que ver con la definición de agentes de red. Hemos visto en los párrafos inciales que como *agente*, *agentes de software* o *agentes de red* se encuentran definiciones todavía muy generales. Intentaremos explicar, no obstante, que esperamos poder hacer en este terreno y qué consecuencias tiene dentro del problema general de la IA.

El Proyecto LOCUS alienta la formación de una comunidad de usuarios ligados por su pertenencia a una determinada entidad, empresa u organización. Se supone que todos ellos elaboran sus documentos añadiendo metadatos a partir de los recursos discutidos en el apartado anterior y que contribuyen solidariamente a la producción de un buscador suficientemente robusto. Aquí nuevamente volcamos la atención en el aspecto del *consenso*. Téngase en cuenta que los contenidos de las etiquetas empleadas en sus documentos son los que luego engrosan los contenedores que sirven para producir preguntas estructuradas. En la medida en que todos los usuarios del entorno contribuyen a enriquecer esta base de datos, aumentan la eficiencia de las preguntas incrementando el éxito en la identificación de posibles respuestas. Así entendido, el acerbo documental anotado que comparte esta comunidad se convierte en una fuente de información altamente estructurada dentro de la cual es posible defi-

nir tareas a ejecutar. Consideremos un ejemplo fácil de entender. En una corporación dotada de varios departamentos las decisiones que cada uno de estos adopta en el uso de su autonomía suelen afectar, a menudo de formas muy diversas, a las restantes unidades de esa organización. Si uno de estos departamentos decide cambiar la persona que actúa como su representante en una comisión parece evidente que habrá que comunicar el cambio al resto de los departamentos que tienen representación en dicha comisión. Si el hecho se refleja en un documento anotado con un sistema de etiquetas como el que representa LOCUS es relativamente fácil transferir el acto de informar a las unidades correspondientes a un agente de red que actúe de forma independiente. Se trata de programar tareas que uno de estos agentes sepa interpretar y realizar con cierta eficiencia. Para ello bastaría identificar ciertos contenidos típicos de los contenedores que los usuarios han ido generando con su uso y proponer una tarea que quede asociada a la aparición de los desencadenantes apropiados. Es obvio que la definición de una tarea no puede ser dejada en principio al propio *criterio* del agente de red que está operando sobre la comunidad en cuestión. Se supone que no tiene control sobre los actos que se derivan del contenido de la base documental y que por tanto carece de intencionalidad e iniciativa. Lo que sí es posible imaginar es un agente capaz de reforzar ciertas tareas e inhibir otras a partir de su interacción con los usuarios del sistema. Si una tarea es aceptada y *agradecida* por la comunidad ésta se fijará en la conducta de nuestro agente, si es rechazada será inhibida. La capacidad de los usuarios para proponer tareas al agente haría el resto.

Esta forma peculiar de automatizar tareas dejando que los usuarios determinen su éxito a través la más pura selección, da idea del nuevo tipo de entorno al que se enfrenta el proyecto de la IA. La consabida idea de la *simulación* se encuentra a la base de juicios de este tipo. Si no podemos encontrar diferencia entre la eficacia de un agente humano – una secretaria, o personal de un archivo, p.ej.– y un sistema informático, a la hora de trabajar con un archivo de documentación –caso que podría darse cotidianamente en una facultad– entraríamos de nuevo en el debate de que con qué criterios podriamos catalogar de inteligente al ordenador o no, vistas ciertas ausencias de diferencia con el humano.

Resulta difícil, no obstante, conjeturar nada acerca de si los agentes de red que pueden definirse a partir de estas intuiciones, responden o no al espíritu genuino de la IA. Nuestra impresión se somete aquí a la duda. El uso de documentos etiquetados con metadatos es una estrategia en cierto modo derivada del fracaso relativo de la IA clásica a la hora de obtener sistemas complejos dotados de habilidades semánticas e intencionalidad. Sin embargo, y como en tantas otras ocasiones, aún es posible hacer del vicio virtud. Los agentes de red pueden desarrollar ciertas capacidades tomando como punto de partida el tipo de tareas que hemos descrito. No creemos que exista un futuro posible para la IA, al margen del uso que comunidades de usuarios más o menos definidas puedan hacer de ciertas herramientas que, como los agentes de red, sólo serían parcialmente autónomos. La IA ha estado por demasiado tiempo enclaustrada en los límites del laboratorio imitando modelos que encuentran su origen en los propios inicios de la Teoría de la Computación. Nuestros modelos no pueden seguir siendo concebidos como auténticos *cerebros en una bañera* si queremos que realmente tengan algún porvenir. Por eso creo que es bueno confiar en agentes limitados en principio pero capaces de mostrarse útiles a quienes les confían ciertas tareas. El

tiempo y la presión del uso dirá en qué dirección son capaces de mejorar y cobrar independencia. La IA del futuro surgirá posiblemente del intercambio efectivo entre comunidades numerosas de usuarios dispuestos a crear documentos fuertemente estructurados y modestos agentes previstos en principio como meros recursos para aligerar pesadas y rutinarias tareas derivadas del uso e intercambio de esa información. Lo que tenga que surgir de todo esto está aún por ver, siendo nuestra responsabilidad participar o no en ello.

Referencias

[1] Naylor, Chris (1983): *Build your own expert system*. Traducción castellana *Construya su propio sistema experto* de G. Fernández, Editado por Díaz de Santos S.A., Madrid, 1986.

[2] Symeonidis, Andreas L. and Mitkas, Pericles A. (2005): *Agent intelligence throught data mining*, Springer Science+Business Media, Inc., 233 Spring Street, New York, 2005.

[3] Mc Carthy, John (1990a): "Some Experts System Need Common sense" localizable en *Fourth International Syposium of Knowledge Engineering: Technical Sessions,* en el apartado de "Plenary Lectures", celebrado en Barcelona del 7 al 10 de Mayo de 1990. Actas publicadas por Rank Xerox.
Veasé también referencia [1], págs. 205-241.

[4] Wooldridge, Michael and Jennings, Nicholas R. (1995): "Intelligent Agents: Theory and Practice", *Knowledge Engineering Review*, 10 (2), págs. 115-152.

[5] Wooldridge, Michael (1999): *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, chapter 1, pages 27-78, MIT Press, Cambridge, MA, USA.

[6] Obitko, Marek and Marík, Vladimír (2003): "Maping between Ontologies in Agent Communication", localizable en *Multi-agent System and Applications III: 3rd International Central and Eastern European Conference On Multi-Agent Systems,* CEEMAS 2003, Prague, Czech Republic, June 2003, Proceedings, págs. 191-203.

[7] W3C (2004): OWL Web Ontology Language Reference. http://www.w3.org/TR/owl-ref

[8] Huang Jinsang, Zavala Gutiérrez Rosa Laura, Mendoza García Benito, Huhns Michael N. (2005): "Reconciling Agent Ontologies for Web Service Applications" en *Multiagent System Technologies: Third German Conference,*MATES 2005, Koblenz, Germany, September 2005, Proceedings, págs. 106-118, Springer-Verlag Berlín, Heildelberg, 2005.

[9] JWLN: Java World Net Library (2003) – JWNL 1.3. http://sourceforge.net/projects/jwordnet

[10] Veansé las posiciones venidas del SL hacia el uso exclusivo del formato .doc en Stallman, Richard (2002). "Podemos acabar con los archivos adjuntos en Word", en http://www.gnu.org/philosophy/no-word-attachments.es.htm

[11] Johnson, Steve (1997); *Interfaz culture: How new Technology transforms the way we create and comunicate,* Harper, San francisco.

[12] Nawarecki Edward, Dobrowolski Grzegorz, Ciszewski Stanislaw, and Kisiel-Dorohinicki Marek (2003): "Ontology of Cooperating Agents by Means of Knowledge Components" en Marik V, et al. (eds): CEEMAS 2003, LNAI 2691, pp.180-190, Springer-Verlag Berlín, Heildelberg 2003.

[13] Interés que encontramos por ejemplo en González Julio Abascal, García Peñalvo, Francisco José y Gil González Ana Belén (eds) (2001): *Interacción' 2001: 2º Congreso Internacional de Interacción Persona-Ordenador*, celebrado los días 16, 17 y 18 de Mayo de 2001 en Salamanca. Actas de las Ediciones de la Universidad de Salamanca, Mayo del 2001.

Una aproximación incremental para adquisición y modelado de conocimiento sobre diagnosis en medicina

M. Taboada[1], J. Mira[2] y J. Des[3]

[1] Dpto. de Electrónica e Computación, Universidad de Santiago de Compostela.
15782 Santiago de Compostela, Spain.
chus@dec.usc.es
http://aiff.usc.es/ elchus/
[2] Dpto. de Inteligencia Artificial, UNED.
28040 Madrid, Spain.
jmira@dia.uned.es
http://www.ia.uned.es/personal/jmira///
[3] Servicio de Oftalmología, Hospital Comarcal *Dr. Julián García.*
27400 Monforte de Lemos, Spain.
eljjdes@telefonica.net

Resumen A partir de la mitad de los 90, la comunidad médica ha empezado a considerar la diagnosis médica como una ciencia que, como tal, comprende la elicitación de las teorías que explican el comportamiento de los expertos cuando llevan a cabo dicha tarea. Se reconoce así que es necesario desarrollar una metodología que guíe el proceso de razonamiento clínico durante la diagnosis médica. Este trabajo proporciona un entorno metodológico que sistematiza el proceso de elicitación de conocimiento a partir de documentos textuales. El entorno proporciona un número limitado de etapas y actividades, en las que se va transformando gradualmente el conocimiento textual en conocimiento estructurado. Además, cada etapa preserva la traza de las transformaciones realizadas, constituyendo una excelente fuente de documentación de todo el proceso de elicitación.

Palabras clave: modelado del conocimiento, ontologías, guías de práctica clínica, diagnosis en medicina

1. Introducción

La diagnosis médica es una actividad rutinaria en la práctica clínica. Cada vez que un paciente consulta a un médico, éste realiza automáticamente un diagnóstico. Por ejemplo, si un paciente le relata que tiene el *Ojo Rojo* y que ha apreciado una disminución de su agudeza visual, entonces el médico empieza a pensar en las posibles causas que pueden haber provocado dicha alteración. Dependiendo de las destrezas del experto clínico, la explicación a los síntomas del paciente puede ser diferente debido, principalmente, al tipo de razonamiento llevado a

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 79-90, 2006.

cabo y al grado de conocimiento sobre dicha dolencia. En general, a mayor experiencia, la calidad del razonamiento mejora y disminuye el tiempo necesario para alcanzar el diagnóstico.

Durante décadas, se ha considerado la diagnosis médica como un *arte* y no como una ciencia. Esto ha repercutido negativamente sobre la metodología de enseñanza aplicada en este dominio, basada en la memorización del conocimiento que caracteriza las enfermedades, en la observación de la forma de trabajo *in situ* de los especialistas y la contrastación de estas observaciones con el conocimiento adquirido a través de los libros. Como resultado de este procedimiento de aprendizaje, se origina idiosincrasia en los razonamientos que los médicos realizan durante la diagnosis, dificultando la estandarización de la práctica clínica. A partir de la mitad de los 90, la comunidad médica ha empezado a reconocer que las tareas cognitivas que se realizan durante la labor clínica constituyen por sí mismas un bloque de conocimiento específico y separado del conocimiento que caracteriza las enfermedades [1,2]. De esta manera, se empieza a considerar la diagnosis médica como una ciencia que, como tal, comprende la elicitación de las teorías que explican el comportamiento de los expertos cuando llevan a cabo la diagnosis médica. Como resultado, hay un convencimiento general de que estas teorías pueden ser transmitidas para su aplicación práctica. Dichas teorías comprenderán procedimientos generales y sistemáticos de diagnóstico que deberían ser reflejo de los métodos de razonamiento usados en la práctica clínica. Se reconoce así que es necesario desarrollar una metodología que guíe el proceso de razonamiento clínico durante la diagnosis médica [3].

Por otra parte, ya desde los años setenta, el campo de la Ingeniería del Conocimiento (IC) se ha volcado en tratar de comprender y simular en el ordenador algunas tareas cognitivas. Muchos de los primeros sistemas expertos de los años 70 se implementaron en el dominio de la medicina, principalmente como sistemas de ayuda al diagnóstico y planificación de terapia. Ejemplos son MYCIN [4], MDX [5], INTERNIST [6] y CASNET [7]. Sin embargo, los estudios sobre los aspectos conceptuales y formales subyacentes a la diagnosis no llegaron hasta mediados de los 80 [8,9,10,11]. Actualmente, no hay una única forma de caracterizar los sistemas de diagnóstico. En general, éstos se pueden clasificar atendiendo a diferentes dimensiones [12]: al tipo de modelo sobre el que razonan (modelos de comportamiento normal y/o anormal), al tipo de inferencia que hacen (abductiva frente a basada en consistencia), al tipo de solución proporcionada (fallo único frente a fallo múltiple), a la semántica de los diagnósticos de salida (diagnóstico etiológico frente a diagnóstico fisiopatológico), etc. La conclusión de todos estos estudios es que la IC puede proporcionar herramientas para analizar las estructuras conceptuales y de razonamiento del conocimiento usado en la diagnosis médica.

En este trabajo nos centraremos en proporcionar estrategias metodológicas que sistematicen, en la medida de lo posible, el proceso de adquisición y modelado de conocimiento sobre diagnosis médica. Propondremos un conjunto de etapas ordenadas y, en cada una de estas, un conjunto de actividades a realizar. El objetivo último del trabajo es proporcionar un soporte metodológico adecuado

para describir y comprender todas las transformaciones que se llevan a cabo cuando el ingeniero adquiere el conocimiento del experto y lo plasma en un modelo de conocimiento. Dicho soporte puede verse como un medio para crear la documentación sobre el proceso de modelado. Dicha documentación es muy importante, tanto por su valor didáctico como de mantenimiento.

La estructura del artículo es la siguiente. Comenzaremos describiendo brevemente y justificando nuestra propuesta. A continuación, detallaremos, con ejemplos, el conjunto de etapas y actividades que incluye nuestra propuesta. Finalmente, presentaremos las conclusiones de este trabajo.

2. Una propuesta orientada a sistematizar el proceso de adquisición y modelado del conocimiento

El tan parafraseado 'Knowledge acquisition bottleneck'[13] hace referencia a la dificultad que siempre ha presentado la Adquisición del Conocimiento (AC). Para minimizar sus efectos, Allen Newell [14] introdujo el *Nivel del Conocimiento*, consiguiendo así separar la AC de su representación simbólica. Posteriormente, surgieron un conjunto de metodologías basadas en esta noción (tales como CommonKADS [15]) que, hoy en día, sirven como herramientas básicas de ayuda para modelar el comportamiento de resolución de problemas. No obstante, el desarrollo de sistemas reales de ayuda al diagnóstico sigue siendo, a día de hoy, una tarea nada trivial. La complejidad sigue planteándose en la etapa de la Adquisición del Conocimiento (AC) debido, principalmente, a que el soporte inicial del conocimiento médico es el lenguaje natural. Los médicos describen su conocimiento en forma de texto plano o parcialmente estructurado, en la guías de práctica clínica, protocolos, explicaciones, historias clínicas, libros de texto, etc. Sin embargo, los ingenieros de conocimiento no son capaces de comprender la semántica de estos textos. Además, en muchos casos, el conocimiento está *implícito* en los textos, de forma que únicamente es claro y transparente para los médicos que lo describen. Por ello, durante el modelado de conocimiento se requiere hacer explícito todo el conocimiento implícito en los textos.

La metodología CommonKADS se centra en la obtención de un modelo de pericia que describa adecuadamente la tarea llevada a cabo por el experto del dominio. Dichos modelos son formulados, en la mayoría de los casos, por el ingeniero del conocimiento apoyado con entrevistas a los expertos del dominio. Podemos decir, pues, que CommonKADS está centrada en la obtención de un modelo de pericia compacto. Una alternativa complementaria que proponemos en este trabajo es re-estructurar el proceso de construcción del modelo de pericia propuesto por CommonKADS con el fin de mantener la traza de las transformaciones durante dicha construcción. Es decir, proponemos establecer correspondencias (*mappings*) explícitas entre los textos médicos originales y el modelo de pericia, almacenando dichas correspondencias de forma estructurada. Al obligar al ingeniero del conocimiento a mantener la traza de las transformaciones reali-zadas, el proceso de modelado del conocimiento se vuelve más gradual. Si además proporcionamos un conjunto de etapas explícitamente definidas, que

guíen y limiten el número de transformaciones a realizar en cada etapa, el modelado de conocimiento se volverá más sistemático y transparente. Con un número limi-tado de transformaciones en cada etapa, el riesgo de falta de conocimiento y ambigüedades en el modelo final se reduce, facilitándose las subsecuentes tareas de verificación, validación y mantenimiento. Todas estas características son importantes principalmente cuando se requiere fusionar conocimiento de diferentes fuentes (guías de práctica clínica, protocolos, descripciones de los expertos, historias clínicas, libros de texto, etc.) y adaptarlo a un entorno con condicionantes particulares, tal y como son los entornos clínicos. La figura 1 muestra este proceso de modelado en líneas generales. El resultado final de este proceso no sólo será un modelo de pericia del dominio en cuestión, sino también un conjunto de correspondencias entre descripciones médicas textuales y componentes de la IC.

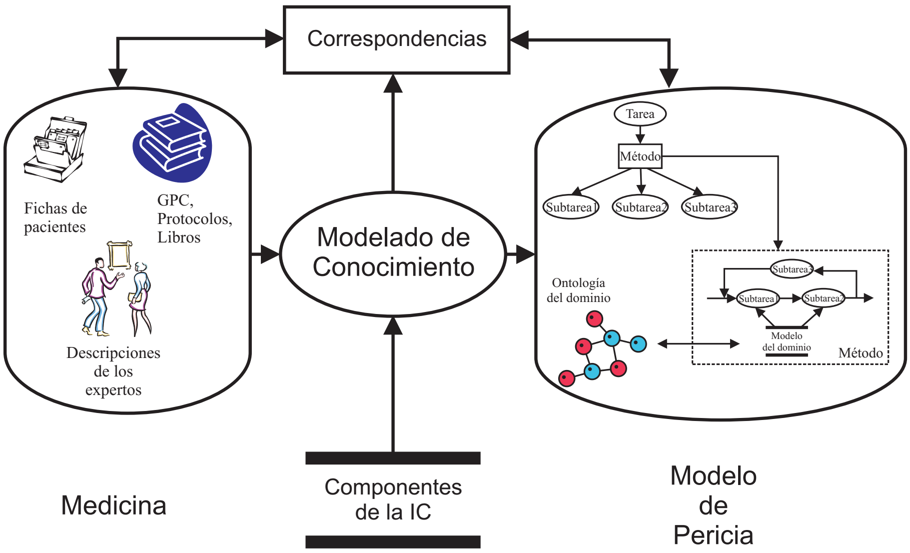


Figura 1. Modelado de conocimiento preservando la traza de las transformaciones.

3. Etapas en la construcción del modelo de conocimiento

Esta sección presenta un conjunto de estrategias para la construcción de un modelo de conocimiento en el dominio de la diagnosis médica. Este proceso incluye tres etapas principales y un conjunto ordenado de actividades a llevar a cabo en cada etapa:

1. *Identificación de las piezas de conocimiento.* En esta etapa, se especifican todas las piezas que se manejarán durante el proceso de modelado, incluyendo

tanto las piezas del conocimiento médico a modelar como los componentes de la IC que se ensamblarán para construir el modelo de conocimiento (Fig. 2).

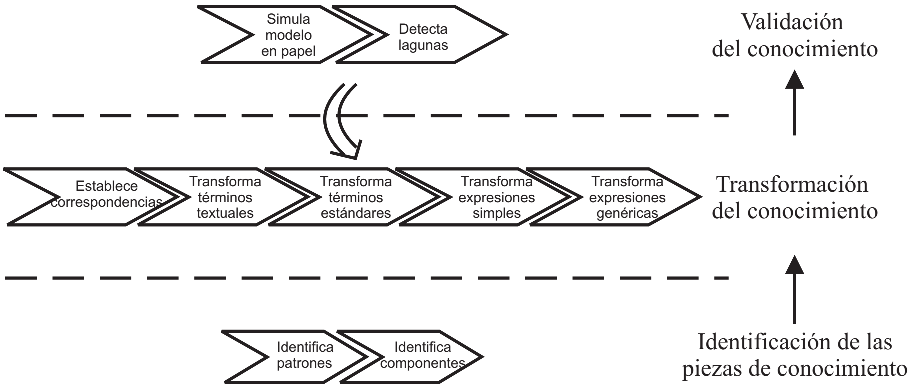


Figura 2. Principales etapas y actividades en la construcción del modelo de conocimiento.

a) *Identificación de patrones de conocimiento.* Esta actividad consiste en analizar las fuentes del conocimiento disponibles e identificar los patrones del conocimiento médico que se requieren modelar. Como ejemplo hemos elegido la documentación textual proporcionada por una guía de práctica clínica en el dominio de la oftalmología [4], debido al valor que, hoy en día, tienen dichas guías como herramientas de apoyo a la estandarización y mejora de la calidad asistencial. En este documentación, hemos identificamos algunos de los patrones de conocimiento que pasamos a detallar a continuación:

- *Sustantivos aislados* y *sustantivos acompañados* de adjetivos. Ejemplos son *diagnosis*, *cause*, *visual function* o *discharge*.
- Expresiones *difusas* o *vagas* que reflejan el estado de las variables clínicas que se pueden medir o cuantificar. Por ejemplo, una expresión como *disminución de la agudeza visual* o *desarrollo rápido de conjuntivitis hiperpurulenta grave*.
- *Expresiones lingüísticas genéricas*, incluyendo
    - Verbos que describen estructuras o partes. Por ejemplo, en la expresión *La población de pacientes incluye individuos de todas las edades que presenten síntomas sugestivos de conjuntivitis ,...* el verbo *incluye* hace referencia al conjunto de pacientes a los que va dirigida la guía de práctica clínica.

[4] http://www.aao.org/aao/education/library/ppp

- Verbos que describen acciones médicas. Por ejemplo, los objetivos relatados en la guía de práctica clínica incluyen acciones clínicas, tales como *diagnosticar* o *aplicar tratamiento*.
- Decisiones o causalidades del tipo 'si...', 'pero...', 'debería ...', etc. Un ejemplo es la expresión *Preguntas sobre los siguientes elementos de la historia del paciente pueden elicitar información útil:* ..., que expresa una indicación sobre qué información recopilar durante el interrogatorio del paciente.

b) *Identificación de los componentes a reutilizar.* Esta actividad consiste en revisar los componentes de modelado, tales como modelos de tareas estereotípicas, ontologías, etc., que han sido previamente diseñados por otros ingenieros de conocimiento. Estos componentes pueden existir bien como componentes conceptuales o bien como componentes software. De esta manera, la construcción del modelo de conocimiento puede verse como una actividad consistente en ensamblar dichos componentes [16]. En el nuestra aplicación, hemos utilizado:
   - La librería de Benjamins [12] porque proporciona una recopilación de métodos de resolución de problemas (MRPs) en el dominio del diagnóstico. Un MRP especifica un patrón de razonamiento independiente del dominio, por lo que puede ser reutilizado en diferentes aplicaciones. El entorno propuesto por Benjamins caracteriza los MRPs mediante criterios de idoneidad. Dichos criterios se pueden ver como una herramienta que facilita la elección de un MRP durante la construcción de un modelo concreto de diagnóstico.
   - El Metathesaurus de UMLS [17] como una fuente de terminología unificada en el dominio médico.
   - La Red Semántica de UMLS como una ontología del dominio médico.

2. *Transformación del conocimiento textual en componentes de conocimiento.* En esta etapa es donde propiamente se establecen las correspondencias explícitas entre los diferentes tipos de patrones y piezas del conocimiento, paralelamente a la obtención del modelo de pericia (Véase nuevamente la figura 2).
   a) *Establecer correspondencias entre el conocimiento textual.* Como los ingenieros del conocimiento no entienden la semántica de los documentos médicos, la primera actividad que proponemos en esta etapa es la revisión de los documentos por parte de los expertos clínicos con el fin de agrupar y relacionar partes textuales no contiguas que hacen referencia al mismo conocimiento. Por ejemplo, en la figura 3 se han agrupado diferentes partes del texto que se refieren al mismo item o que detallan, más en profundidad, otros párrafos.
   b) *Transformar términos textuales en términos estándares.* Los sustantivos incluidos en los textos deben extraerse y sustituirse por términos estándares. A esta actividad se le conoce comúnmente como *Adquisición de terminología médica estándar*. El uso de los recursos proporcionados por algún sistema de terminología médica unificada puede ayudar a semiautomatizar dicho proceso. Por ejemplo, en nuestro caso de estudio,

hemos utilizado las utilidades que proporciona el servidor del UMLS[5]. Así, hemos podido establecer correspondencias directas entre términos textuales como *diagnosis* o *terapia* y sus correspondientes términos estándar (Véase Cuadro 1). Además, durante el desarrollo de esta actividad hay que tener en cuenta que los sustantivos deben extraerse fuera de su contexto sin pérdida de conocimiento. Esta transformación puede conllevar un proceso de *Refinamiento de los términos*. Por ejemplo, el término *secreción (discharge)* debería refinarse en *secreción ocular*. Por otra parte, los términos demasiado especializados (y, por tanto, no presentes en la terminología médica utilizada) deben sustituirse por términos más generales (*Generalización de términos*) y representar las características especiales como atributos de los términos generales (en la siguiente actividad). Por ejemplo, *leve secreción mucosa* es un término tan especializado que no se encuentra, como tal, en el Metathesaurus del UMLS. Sin embargo, sí se encuentra el término más general *secreción ocular mucosa*.

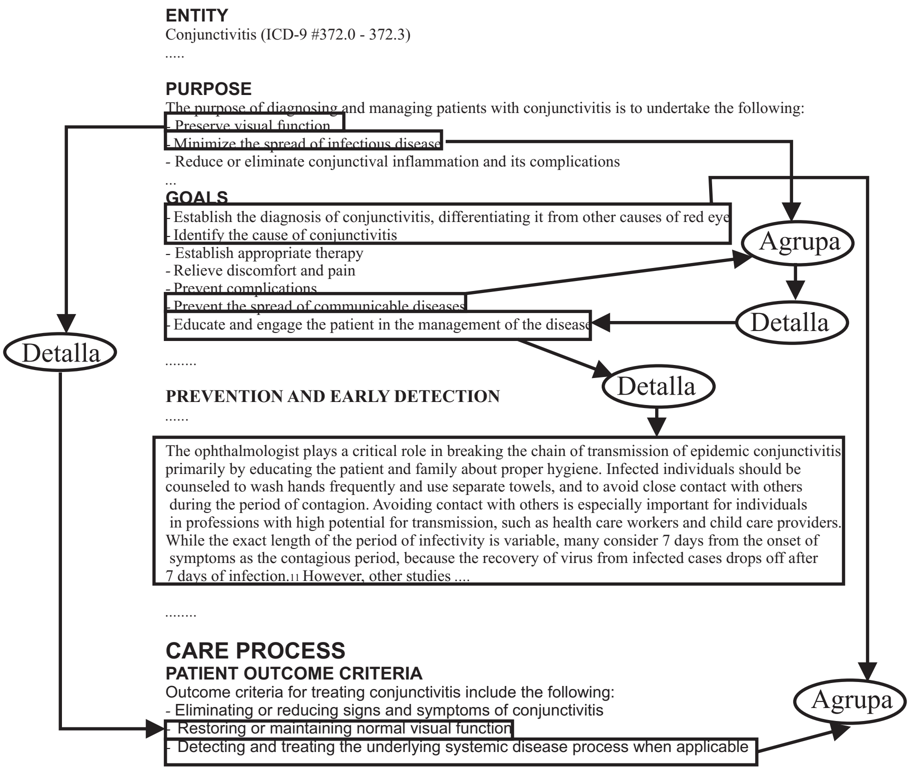


Figura 3. Correspondencias marcadas entre contenidos de una guía de práctica clínica sobre conjuntivitis publicada por la Academia Americana de Oftalmología.

[5] http://www.nlm.nih.gov

Cuadro 1. Ejemplos de transformación de términos textuales

Término Textual	**Tipo de Correspondencia**	**Término Estándar**
diagnosis	Equivale a	Diagnosis (CUI: C0011900)
therapy	Equivale a	Therapeutic procedure (CUI: C0087111)
discharge	Se refina en	Discharge from eye (CUI: C0423006)
mild mucous discharge	Se generaliza en	EYE DISCHARGE, MUCOID (CUI: C0239425)
duration	Equivale a	Duration (CUI: C0449238)

c) *Transformar términos estándares en piezas de conocimiento* (conceptos, propiedades, relaciones entre conceptos, tareas estereotípicas, MRPs, roles de conocimiento, ...). Para llevar a cabo esta actividad, tenemos en cuenta la información proporcionada por la Red Semántica del UMLS. Esta red proporciona una ontología médica que clasifica cada uno de los conceptos del Metathesaurus. Por tanto, los conceptos del Metathesaurus extraídos en la actividad previa tienen asociado un tipo semántico, que nos proporciona una pista sobre la pieza de conocimiento a la que se puede referir. Por ejemplo, *Diagnosis (CUI: C0011900)* es un *Procedimiento diagnóstico* y *Therapeutic procedure (CUI: C0087111)* es un *Procedimiento preventivo y terapéutico*, por lo que pueden modelarse como tareas o MRPs. Nosotros las etiquetamos como tareas estereotípicas. *Discharge from eye (CUI: C0423006)* es un *Signo o Síntoma* y *Duration (CUI: C0449238)* es un *Concepto Temporal*, por lo que se corresponden con conceptos del dominio. Por otra parte, añadimos la propiedad *duration*, entre otras, a los *Signos o Síntomas*.

Cuadro 2. Ejemplos de transformación de términos estándares

Término Estándar	**Tipo de Correspondencia**	**Pieza de conocimiento**
Diagnosis (CUI: C0011900)	Se corresponde con	Tarea de Diagnosis
Therapeutic procedure (CUI: C0087111)	Se corresponde con	Tarea de Valoración
Discharge from eye (CUI: C0423006)	Se corresponde con	Concepto del dominio
Duration (CUI: C0449238)	Se corresponde con Se añade	Concepto del dominio Atributo del dominio

d) *Transformar expresiones simples relativas al estado de las variables clínicas.* En esta actividad principalmente se deciden qué atributos añadir a los conceptos, en función de las expresiones analizadas, asignándoles rangos de valores. Por ejemplo, la expresión *desarrollo rápido de conjuntivitis hiperpurulenta grave* pone de manifiesto la necesidad de añadir

el atributo *Severity (CUI:C0449294)* para graduar los síntomas y el atributo *Onset* para describir la forma en cómo se instauran. Posibles valores para el atributo *Severity* son *Severe (CUI: C0205082)* y los conceptos hermanos en el Metathesaurus: *Mild (CUI: C0205080)* y *Moderate (CUI: C0205081).* Posibles valores para el atributo *Onset* son *Fast (CUI: C0456962)* y los conceptos hermanos en el Metathesaurus: *Gradual (CUI: C0439833)* y *Sudden (CUI: C0439832)*, etc. Como el término hiperpurulenta no existe en el Metathesaurus, lo refinamos en *purulenta*. Además, purulenta es un atributo que describe las secreciones, no la propia enfermedad de la conjuntivitis, por lo que refinamos *conjuntivitis hiperpurulenta grave* en *secreción purulenta grave*. Con relación al *desarrollo rápido*, es necesario refinar la expresión en el tiempo concreto que significa dicho desarrollo. Por ejemplo, 1-2 días.

e) *Transformar expresiones lingüísticas genéricas.* Las expresiones lingüísticas genéricas deberían ser transformadas a componentes de conocimiento. Generalmente esta transformación conlleva refinamiento de conocimiento para poder adaptar las descripciones en lenguaje natural a las características particulares del entorno. Ejemplos de tales expresiones son:
   - Verbos que describen estructuras o partes se pueden transformar en conceptos o relaciones estándar de la ontología del dominio. Por ejemplo, en la expresión *La exploración ocular inicial incluye medida de la agudeza visual, exploración externa y exploración con lámpara de hendidura*, el verbo *incluye* describe las partes de que debe constar una exploración ocular. Es, por tanto, una relación entre componentes de una ontología.
   - Verbos que describen acciones se pueden transformar en métodos de resolución de problemas, inferencias o funciones de transferencia. Por ejemplo, un objetivo relatado en la guía de práctica clínica es *Establecer el diagnóstico de conjuntivitis, diferenciándolo de otras causas de ojo rojo.* En esta expresión, el sustantivo *diagnosis* indica una tarea de la IC (Cuadro 2). Dicho sustantivo está matizado en la expresión mediante el verbo *differentiate* que se corresponde con el concepto del Metathesaurus *Differential Diagnosis (CUI: C0011906).* A su vez, dicho concepto se puede hacer corresponder con el método de la IC conocido como Método principal de diagnóstico [12], que descompone la tarea de diagnóstico en tres subtareas: detección de síntomas, generación de hipótesis y discriminación de hipótesis.
   - Decisiones o causalidades (*si ..., pero ..., debería ser ..*, etc.) se pueden transformar en tareas y/o MRPs. Por ejemplo, la descripción *La exploración externa debería incluir los siguientes elementos: ...*, el verbo *incluye* describe las partes de que debe constar una exploración ocular externa, en condiciones ideales de trabajo. Es, por tanto, una relación entre componentes de una ontología. Pero además, al venir matizado como *debería incluir* también indica que, en la práctica, la lista es demasiado exhaustiva y el médico debe decidir, en cada

momento, qué elementos son los más importantes para el paciente en cuestión. La selección de las partes es dinámica y depende de los datos particulares del paciente. Por tanto, dicha expresión la hemos hecho corresponder con una tarea de valoración (assessment) consistente en recopilar datos adicionales, teniendo en cuenta las hipótesis actuales del diagnóstico así como los costes implicados en la exploración externa.

3. *Validación del conocimiento.* Esta etapa incluye dos actividades a realizar:
   a) *Validar*, tanto como podamos, *las transformaciones de conocimiento* realizadas en etapas anteriores. Una importante técnica, para validar la especificación final del modelo de conocimiento obtenido, es intentar hacer una simulación en papel, a partir de escenarios reales de la diagnosis médica.
   b) *Detección de lagunas de conocimiento.* La validación realizada en la actividad anterior reflejará si el modelo construido permite simular el comportamiento médico deseado, permitiendo detectar si está completo o no. En este último caso, deberemos volver a realizar alguna actividad de las etapas previas. En nuestro ejemplo, nos encontramos que la guía de práctica clínica incluye conocimiento sobre algunas dimensiones caracterizando el diagnóstico de la conjuntivitis, tales como *diagnóstico diferencial* con respecto a otras causas de *Ojo Rojo*, *diagnóstico etiológico* (para identificar la causa de conjuntivitis), *diagnóstico de fallo múltiple*, conocimiento sobre el curso de la enfermedad y sobre los síntomas y signos asociados a cada tipo de patología. Sin embargo, no contempla conocimiento específico sobre los métodos que emplea el médico cuando genera hipótesis diagnósticas a partir de los síntomas y signos del paciente, cuando decide cómo realizar la exploración ocular y cómo evaluar las hipótesis diagnósticas. En nuestro caso, esta información se ha adquirido directamente de entrevistas con el médico especialista.

4. Conclusiones

En este artículo proponemos un entorno metodológico orientado a sistema-tizar el proceso de adquisición y modelado de conocimiento. Nuestro entorno amplía el proceso de construcción de modelos de pericia de CommonKADS [15], al proporcionar un soporte metodológico que ayuda a detectar y documentar todas las transformaciones del conocimiento en formato textual al formato estructurado del modelo de pericia. Al obligar al ingeniero del conocimiento a mantener la traza de las transformaciones realizadas, el proceso de modelado del conocimiento es más gradual. Además proporcionamos un número limitado de transformaciones en cada etapa, por lo que el riesgo de falta de conocimiento y ambigüedades en el modelo final se reduce, facilitándose las subsecuentes tareas de verificación, validación y mantenimiento.

En el artículo hemos presentado una aplicación del entorno a la elicitación y modelado de una guía de práctica clínica. Dicha elicitación se ha hecho en

términos de componentes reutilizables de conocimiento: ontologías del dominio, MRPs, tareas estereotípicas, etc. Sin embargo, esta forma de adquisición de guías de práctica clínica no ha sido la usual en las últimas décadas. Inicialmente, las guías de práctica clínica empezaron a codificarse directamente en alguno de los lenguajes de representación propuestos para tal fin. En [18] se puede encontrar una revisión de las propuestas más importantes. En general, estos lenguajes son muy expresivos, permitiendo describir el contenido de una guía de práctica clínica de forma precisa e inambigüa, en términos de un conjunto de primitivas, tales como acciones y decisiones. Sin embargo, la codificación de una guía directamente en estos lenguajes es una tarea ardua y compleja, y el modelo resultante final es ilegible para los especialistas médicos, dificultando su validación. En estos últimos años, están surgiendo propuestas para describir las guías en un nivel de abstracción mayor. Por ejemplo, en [19] se propone modelar el conocimiento de una guía combinando ontologías, MRPs y primitivas. La diferencia en nuestra propuesta es separar claramente la etapa de modelado de conocimiento de la etapa de representación simbólica. Otra propuesta en la línea de nuestro trabajo la podemos encontrar en [20], donde se ha desarrollado una herramienta para marcar texto libre y transformarlo en alguna de las cuatro categorías predefinidas: componente procedimental, de causalidad, de objetivo y de definición de conceptos. La ventaja de la herramienta es que usa la tecnología XML, con lo que se facilita al usuario el marcado y la fragmentación del texto en modelos de componentes (en la forma de DTDs) interconectados. Sin embargo, no se proporciona al usuario con un conjunto de actividades que guíen las etapas de marcado semántico de texto y transformación a alguno de los cuatro componentes predefinidos, con lo que el proceso de transformación y modelado puede no ser tan obvio para el usuario final. Nuestra propuesta complementa esta metodología al proporcionar un conjunto de etapas claramente diferenciadas durante la transformación y modelado de conocimiento, en un nivel de abstracción más elevado.

Agradecimientos

Este trabajo ha sido financiado por la Secretaria Xeral de Investigacion e Desenvolvemento de la Xunta de Galicia, a través del proyecto de investigación PGIDT01-PXI20608PR.

Referencias

1. Kassirer, J. Teaching Problem-Solving - How are we doing?. The New England Journal of Medicine, **332** (1995), 1507-1509.
2. Sobel, B., Levine M. Medical education, evidence-based medicine and the disqualification of physician-scientists. Experimental Biology and Medicine, **226** (2001), 713-716.
3. Sadegh-Zadeh, K. Fundamentals of clinical methology 4. Diagnosis. Artificial Intelligence in Medicine, **20** (2000), 227-241.
4. Buchanan, B. G. and Shortliffe,E. Rule-based Expert Systems. Addison- Wesley (1984).

5. B. Chandrasekaran and S. Mittal. Deep versus compiled knowledge approaches to diagnostic problem solving. International Journal of Man-Machine Studies, **19** (1983), 425-436.
6. Miller R. A. Internist-1/CADUCEUS: Problems Facing Expert Consultant Programs. Meth. Inform. Med., **23** (1984), 9-14.
7. Weiss, S., Kulikowski, C., Amarel, S. and Safir. A. A Model-Based Method for Computer-Aided Medical Decision-Making. Artificial Intelligence **11** (1978), 145-172.
8. Clancey, W.J. Heuristic classification. Artificial Intelligence, **27** (3) (1985), 289-350.
9. Reggia, J. A., Nau, D. S. and Wang, Y. Diagnostic expert systems based on a set-covering model. International Journal of Man-Machine Studies, **19** (1983), 437-460.
10. Reiter, R. A theory of diagnosis from first principles. Artificial Intelligence, **32** (1987), 57-95.
11. de Kleer, J., Mackworth, A.K. and Reiter, R. Characterizing diagnoses and systems. Artificial Intelligence, **52** (1992), 197-222.
12. Benjamins, V.R. Problem Solving Methods for Diagnosis, PhD thesis, University of Ams-terdam, (1993).
13. Buchanan, B.G., Barstow, R., Bechtal, R., Bemet, J., Clancey, W., Kulikowski, C., Mit-chell, T. and Waterman, D.A. Constructing an expert system. Hayes-Roth, Waterman and Le-nat (eds.). Building Expert Systems. Addison-Wesley, (1983) 127-167.
14. Newell, A. The knowledge level. Artificial Intelligence, 18 (1982) 87-127.
15. Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., Van de Velde, W. and Wielinga, W. Knowledge Engineering and Management, The CommonKADS Methodology. The MIT Press, (1999).
16. Taboada, M., Des, J., Mira, J. and Marín, R. Development of diagnosis systems in medicine with reusable knowledge components. IEEE Intelligent Systems, **16** (2001), 68-73.
17. Lindberg, D., Humphreys, B. and Mc Cray, A. The Unified Medical Language System. Methods of Information in Medicine, **32** (1993), 281-291.
18. de Clerq, P., Blom, J., Korsten, H. and Hasman, A. Approaches for creating computer-interpretable guidelines that facilitates decision support, Artificial Intelligence in Medicine **31** (2004), 1-27.
19. de Clerq, P., Hasman, A., Blom, J. and Korsten, H. The application of ontologies for the development of shareable guidelines, Artificial Intelligence in Medicine **22** (2001), 1-22.
20. Svatek V., Ruzicka M. Step-by-step formalisation of medical guideline content. International Journal of Medical Informatics, **70** (2-3) (2003), 329-335.

Fusión automatizada de ontologías: Aplicación al razonamiento espacial cualitativo

Joaquín Borrego-Díaz y Antonia M. Chávez-González

Departamento de Ciencias de la Computación e Inteligencia Artificial.
E.T.S. Ingeniería Informática-Universidad de Sevilla.
Avda. Reina Mercedes s.n. 41012-Sevilla
{jborrego, tchavez}@us.es

Resumen La evolución de las ontologías es un problema clave en la Integración del Conocimiento, cuya resolución es imprescindible en el proyecto de la Web Semántica. Algunas aproximaciones adolecen de confianza lógica, y otras no son fácilmente mecanizables. En este trabajo proponemos un método para la fusión de ontologías utilizando razonamiento automático. Como ilustración, presentamos una aplicación en el campo del razonamiento espacial cualitativo, fusionando la ontología sobre mereotopología denominada *Cálculo de Conexión de Regiones*, con otra sobre el tamaño relativo de entidades espaciales.

1. Introducción

Durante sus primeros cincuenta años de vida oficial, la Inteligencia Artificial (IA) se ha enfrentado a grandes retos. Uno de ellos, implícitamente planteado en la conferencia de Dartmouth, es la formalización del conocimiento común. Se puede afirmar que éste es uno de los mayores obstáculos para la creación de agentes plenamente racionales. Este problema ha centrado los esfuerzos de especialistas en Representación del Conocimiento durante todo este tiempo y, con especial énfasis, en los últimos años con el proyecto de la Web Semántica (WS). Dicho proyecto tiene como objetivo el procesamiento mecánico y lógicamente fiable de la información contenida en la WWW [11]. La solución adoptada, ya estudiada para otros proyectos de menor envergadura que la WS, consiste en describir formalmente los elementos del universo de los que trata la información. Es decir, se basa en la construcción de *ontologías*. Usando ontologías se puede referenciar la información para dotarla de un *significado* procesable por máquinas.

Sin embargo, no es suficiente la mera construcción de una ontología. Es evidente que la información no será estática. Las ontologías deben mantenerse como cualquier otra componente de los sistemas. La reutilización del conocimiento requiere que las ontologías sean extendidas, refinadas o integradas [28]. La *evolución* de las ontologías se convierte así en uno de los temas críticos en la WS, implicando tanto problemas de Representación del Conocimiento como de Procesamiento Inteligente de la Información.

Uno de los subproblemas implicados es el de la *integración* de ontologías. La aceptación de lenguajes de representación de ontologías como OWL ha facilitado la proliferación de las mismas. Surge, por tanto, la necesidad de relacionarlas

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 91-102, 2006.

entre sí, para aprovechar conjuntamente el conocimiento aportado por diferentes ontologías. Básicamente, existen tres tipos de reconciliación del conocimiento:

1. la *fusión* (*ontology merging*), que produce una nueva ontología a partir de la mezcla de las ontologías iniciales,
2. la *alineación*, que establece relaciones entre los elementos de las dos ontologías, y
3. la *integración*, que no las une en una sola ontología, sólo establece mecanismos para utilizarlas conjuntamente (completas o en parte).

Para el análisis del proceso de fusión, desde el punto de vista de la lógica computacional, parece necesaria la adopción de nuevas nociones lógicas que estimen la fiabilidad del dicho proceso. Sin embargo, existen serias dificultades. La fusión es difícil de automatizar y se necesita la interacción contínua con el usuario, o usar aproximaciones lingüísticas. De ahí que se desconozca, en general, el efecto de la fusión sobre el razonamiento automático con ontologías [4].

El objetivo de este artículo es proponer una definición formal de fusión entre ontologías con fiabilidad lógica, siguiendo ideas esbozadas en [4], y sugeridas por métodos de depuración de bases de conocimiento referenciadas con ontologías [2,12]. Asimismo, presentamos un método, asistido por Sistemas de Razonamiento Automático (SRA), para obtener dicha fusión. El método se basa en la extensión de ideas sobre extensiones ontológicas robustas presentadas en [8,9].

El método se ilustra fusionando dos ontologías sobre relaciones espaciales. Este tipo de ontologías son muy útiles para el Razonamiento Espacial Cualitativo (REC). El caso del REC es paradigmático, fundamentalmente, por dos motivos. En primer lugar, en el REC es crucial que el razonamiento sobre el espacio no se contamine con información no deducible de la propia ontología (información que proviene de la intuición). De este modo se comprende mejor la potencia deductiva de ésta. El uso de SRA previene esta contaminación, y nos asegura que los resultados obtenidos se siguen lógicamente de la ontología. En segundo lugar, es común en el REC diseñar una ontología *ad hoc* para cada tipo de problemas, donde sólo se representan aspectos parciales del espacio-tiempo. Por tanto, es necesario combinar adecuadamente diversas ontologías en casos complejos como, por ejemplo, en los Sistemas de Información Geográfica (SIG).

Concretamente, fusionaremos dos ontologías bien conocidas. La primera es el Cálculo de Conexión de Regiones (RCC) [15]. La segunda es la micro-ontología más sencilla sobre el tamaño relativo de entidades espaciales, y que denominaremos SIZE. La necesidad de esta fusión surge de la percepción humana. Por ejemplo, se sabe que algunas relaciones topológicas como *A está incluido propiamente en B* sólo son posibles si los objetos implicados poseen tamaños relativos adecuados. De ahí que sea muy interesante el manejo conjunto de RCC con SIZE. Como son ontologías esencialmente diferentes (la *articulación* de las dos ontologías es muy simple), parece adecuado fusionarlas.

La estructura del artículo es como sigue. En la siguiente sección se presentan las dos teorías espaciales a fusionar, describiendo sus propiedades fundamentales. A continuación se analizan posibles soluciones al problema de la fusión (sección 3). En la sección 4 formalizamos la idea de fusión retículo-categórica y se presenta

un método para obtener tales fusiones. Este método se aplica al caso de las ontologías que nos ocupan (en 4.1). La última sección está dedicada a analizar trabajos relacionados y a comentar el trabajo futuro.

El método que proponemos está asistido por dos Sistemas de Razonamiento Automático (SRA), un buscador de modelos (para explorar extensiones de ontologías) y un demostrador automático (para certificar hechos acerca de esas extensiones). En este artículo usamos dos sistemas complementarios programados por W. MCune, OTTER y MACE4 (`http://www-unix-mcs.anl.gov`). El primero de éstos es un demostrador automático, basado en resolución, que permite gran autonomía. El segundo, MACE4, es un buscador de modelos basado en el algoritmo de Davis-Putnam-Loveland-Longemann para decidir la satisfactibilidad. El uso combinado de los dos tipos de sistemas es muy útil en la clasificación/axiomatización de diversas teorías (véanse por ejemplo [24,14]).

2. Dos ontologías para el razonamiento espacial

En este artículo se trabajará con dos microteorías sobre REC ampliamente estudiadas. La primera de éstas es el *Cálculo de Conexión de Regiones* (RCC) [15], una ontología sobre relaciones mereotopológicas. Para RCC, las entidades espaciales son conjuntos regulares no vacíos[1]. La relación básica es la conexión, $C(x, y)$, con significado intuitivo: *las clausuras de x e y se cortan.* La axiomatización de RCC está formada por dos axiomas básicos sobre C,

$$A_1 := \forall x[C(x, x)], \qquad A_2 := \forall x, y[C(x, y) \rightarrow C(y, x)]$$

junto con un conjunto de fórmulas que definen las restantes relaciones espaciales (véase la fig. 1)[2]. RCC prueba que dichas relaciones componen el retículo que se muestra en la figura 6 [12]. Los modelos topológicos de RCC han sido investigados por N.M. Gotts en [19], aunque también es posible estudiar la teoría mediante modelos de tipo algebraico [30]. El razonamiento (espacio)temporal con RCC ha sido ampliamente estudiado [33,27].

El conjunto (exhaustivo) de relaciones disjuntas dos a dos de la figura 2 se denota por RCC8, y RCC5 es $\{DR, PO, PP, PPi, EQ\}$. Se ha constatado empíricamente que RCC8 es más adecuado que RCC5 para representar las relaciones topológicas percibidas por los humanos [22]. Podríamos decir que RCC8 habla de relaciones *sensibles* a las fronteras de las regiones mientras que RCC5 no las tiene en cuenta. RCC8, como cálculo, ha sido profundamente estudiado por J.R. Renz [27] entre otros, y es utilizada en SIG actuales y en bases de datos espaciales. La propia teoría RCC ha sido usada como meta-ontología para analizar anomalías en ontologías [12], y como herramienta para repararlas [3].

A pesar de las propiedades de RCC, esta teoría es claramente insuficiente para razonar sobre otros aspectos del espacio. Surge, por tanto, la necesidad de enriquecer RCC con otras características para razonar, por ejemplo, sobre distancia cualitativa, orientación, convexidad o tiempo (véanse [15,33,18]). Incluso

[1] Un conjunto de un espacio topológico es regular si coincide con el interior de su clausura.

[2] La teoría tiene otros axiomas [15], pero no serán usados en este trabajo.

$DC(x, y) \leftrightarrow \neg C(x, y)$	(x está desconectado de y)
$P(x, y) \leftrightarrow \forall z[C(z, x) \rightarrow C(z, y)]$	(x es parte de y)
$PP(x, y) \leftrightarrow P(x, y) \wedge \neg P(y, x)$	(x es parte propia de y)
$EQ(x, y) \leftrightarrow P(x, y) \wedge P(y, x)$	(x es idéntico a y)
$O(x, y) \leftrightarrow \exists z[P(z, x) \wedge P(z, y)]$	(x e y se solapan)
$DR(x, y) \leftrightarrow \neg O(x, y)$	(x y son discretos)
$PO(x, y) \leftrightarrow O(x, y) \wedge \neg P(x, y) \wedge \neg P(y, x)$	(x e y se solapan parcialmente)
$EC(x, y) \leftrightarrow C(x, y) \wedge \neg O(x, y)$	(x e y están exter. conectados)
$TPP(x, y) \leftrightarrow PP(x, y) \wedge \exists z[EC(z, x) \wedge EC(z, y)]$	(x es parte prop. tang. de y)
$NTPP(x, y) \leftrightarrow PP(x, y) \wedge \neg\exists z[EC(z, x) \wedge EC(z, y)]$	(x es parte propia no tang. de y)

Figura 1. Axiomas de RCC

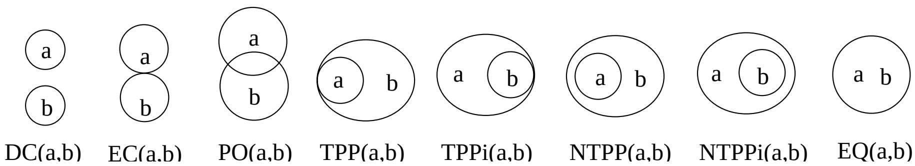


Figura 2. Las relaciones de RCC8

el diseño de lenguajes para representar con metadatos información espacial [16] sugiere la extensión de RCC.

La segunda ontología, que denominaremos SIZE, es la micro-ontología natural sobre el tamaño relativo de entidades espaciales. SIZE ya ha sido usada conjuntamente con RCC en [18], donde A. Gerevini y J.R. Renz extienden el estudio de problemas de satisfacción de restricciones en RCC8.

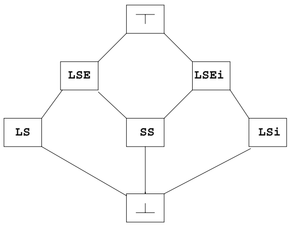
Las relaciones de la ontología SIZE son $LS(x, y)$ (*x tiene menor tamaño que y*) y su inversa $LSi(x, y)$, la relación LSE (*x tiene menor o igual tamaño que que y*) y su inversa $LSEi(x, y)$, y $SS(x, y)$ (*x e y tienen el mismo tamaño*).

Algunos de los axiomas de SIZE se describen en la figura 3 (derecha). SIZE prueba que sus relaciones componen el retículo de la figura 3 (izquierda). Existen otras ontologías sobre el tamaño relativo, véase [21]. SIZE no trata del tamaño cualitativo de las regiones, para el que se utilizan adjetivos (predicados 1-arios), que dependen del contexto donde se usan. Véase también en [21].

2.1. Articulación de RCC y SIZE

La articulación asociada a la relación entre dos ontologías es una ontología intermedia que se puede sintetizar a partir de la relación existente entre los términos de dos ontologías [6]. En nuestro caso, la relación entre las ontologías viene dada por las relaciones descritas en la figura 4 (extraido de [18]). La articulación resultante es simple, prácticamente subretículo de RCC (veáse la fig. 5, izq.).

En este trabajo describimos cómo fusionar las dos ontologías en una nueva, respetando la articulación, y satisfaciendo cierto criterio de minimalidad. Nótese, finalmente, que SIZE es independiente de la dimensionalidad del espacio, es decir, es válida por ejemplo para el razonamiento con intervalos temporales. Así, de



$$\forall x,y(LS(x,y)\vee LSi(x,y)\vee SS(x,y))$$
$$\forall x,y(LS(x,y)\rightarrow\neg(LSi(x,y)\vee SS(x,y))$$
$$\forall x,y(LSE(x,y)\leftrightarrow LS(x,y)\vee SS(x,y))$$
$$\forall x,y(SS(x,y)\leftrightarrow SS(y,x))$$
$$\forall x,y(LS(x,y)\leftrightarrow LSi(y,x))$$

Figura 3. Algunos axiomas de SIZE y el retículo asociado

$$E' := \begin{cases} TPP \sqsubseteq LS & EQ \sqsubseteq SS \\ NTPP \sqsubseteq LS & SS \sqsubseteq DC \sqcup EC \sqcup PO \sqcup EQ \\ TPPi \sqsubseteq LSi & LSi \sqsubseteq DC \sqcup EC \sqcup PO \sqcup TPPi \sqcup NTPPi \\ NTPP \sqsubseteq LSi & LS \sqsubseteq DC \sqcup EC \sqcup PO \sqcup TPP \sqcup NTPP \end{cases}$$

Figura 4. Relación entre RCC y SIZE

manera similar, se podría fusionar SIZE y la ontología clásica sobre intervalos temporales de J. Allen [1].

3. Evolución de ontologías espaciales

Podemos considerar la evolución de una ontología como una sucesión de extensiones y revisiones. Estos procesos están íntimamente ligados. Toda extensión es de hecho una revisión del conocimiento. Sin embargo, es evidente que la revisión inducida por una extensión es una tarea fácil de llevar a cabo con un editor de ontologías (usando p.e. PROTÉGÉ `http://protege.stanford.edu`).

La revisión de una ontología podría ser considerada como una tarea de revisión de creencias, ya estudiada en IA. Sin embargo, este caso es significativamente distinto. Una revisión puede provocar una reinterpretación de los elementos que la ontología representa. Por ejemplo, la propia teoría RCC es, de hecho, una revisión de este tipo de la teoría mereológica de Clarke [13], (que, a su vez, es una revisión de la mereología de Whitehead [32]). Otro aspecto a tener en cuenta es que en la WS, la evolución debe respetar ciertos principios básicos de *compatibilidad hacia atrás* [20]. Es decir, la fusión debe preservar ciertas características fundamentales de las ontologías iniciales.

Existen básicamente dos posibles formas de fusionar dos ontologías, O_1 y O_2, mediante extensiones. La primera, que no es útil en nuestro caso, consiste en extender reiteradamente O_1 definiendo los términos de O_2 (con el lenguaje de O_1). Esto produciría una extensión conservativa de O_1 [5]. Como la obtención de tales definiciones no es posible en general, se podría pensar en la *inserción ontológica* de los términos de O_2 en O_1. Para que la extensión tenga las propiedades adecuadas, es necesario diseñar axiomas que relacionen los términos de las dos ontologías, para preservar ciertas propiedades básicas [8]. Dicha inserción puede implicar una reinterpretación ontológica de algunos elementos de la fusión [12].

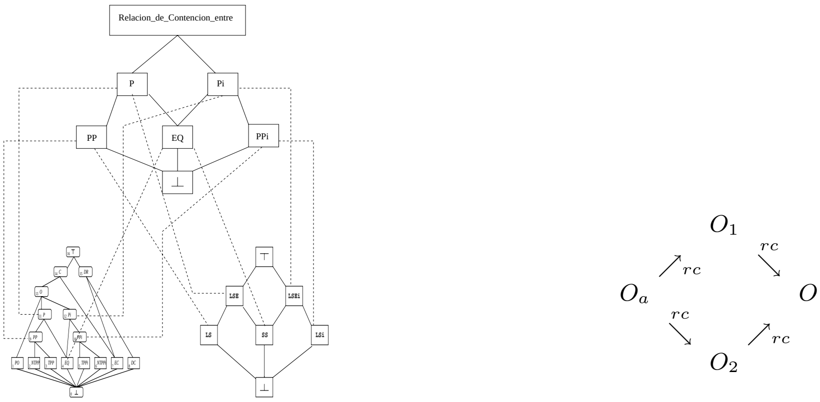


Figura 5. Articulación de RCC y SIZE y diagrama de fusión retículo categórica (derecha)

4. Fusión de ontologías retículo-categóricas

Un método intermedio entre las dos opciones anteriormente comentadas es explorado en [8,9]. La idea se basa en debilitar los principios bajo los que se rige la *metodología definicional* para el diseño de ontologías [7], de forma que se puedan insertar nuevos elementos y sólo se requiere que la extensión tenga cierto grado de categoricidad. Básicamente, sólo se exige que la extensión contenga suficiente información para deducir que el conjunto de conceptos posee la estructura pretendida por el diseñador. A continuación describimos brevemente esta idea (véanse [8,9] para más detalles).

Se supondrá a partir de ahora que el conjunto de conceptos de la ontología estudiada tiene estructura de retículo. Esta restricción es meramente formal; de hecho, muchos métodos de extracción y análisis de ontologías trabajan con retículos de conceptos (como el Análisis Formal de Conceptos [17]). Incluso, es usual en el diseño de ontologías de alto nivel (véase por ejemplo [29]). Aunque la técnica está pensada para ontologías descritas en Lógicas de la Descripción [10] (base lógica para el lenguaje OWL, `http://www.w3.org/TR/owl-features/`), es válida para teorías de primer orden en general.

Si consideramos una teoría T sobre el conjunto de conceptos $\mathcal{C} = \{C_1, \ldots, C_n\}$, y M es modelo de T, denotaremos por $R(M,T)$ el retículo formado por las interpretaciones de los conceptos de $\mathcal{C}$ en M. Una teoría T es retículo-categórica si ese retículo es único (independiente del modelo) salvo isomorfismo. Es evidente que esa noción es más débil que la de categoricidad (de hecho, RCC es retículo categórica [12] pero no es categórica).

Precisemos un poco esta noción. Existe una relación natural entre las propiedades de $R(M,T)$ y la propia teoría T, que hacen muy útil una descripción lógica de dicho retículo para trabajar con los conceptos de T [8]. Una descripción ecuacional E (en el lenguaje de retículos)[3] de $R(M,T)$ se denomina *esqueleto.*

[3] Para mayor claridad, los esqueletos se describirán en Lógica Descriptiva.

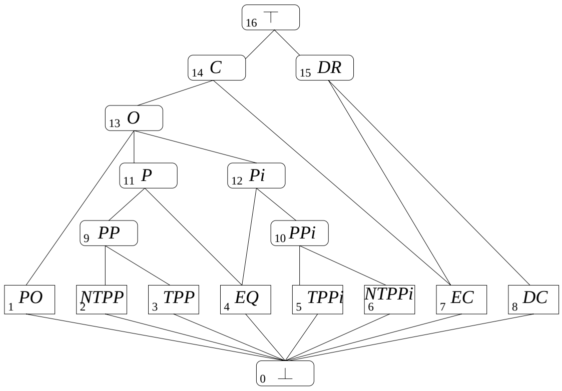


Figura 6. El retículo de las relaciones espaciales de RCC

$$\begin{array}{rlrlrl}
\top \equiv & C \sqcup D & PO \sqsubseteq & \neg P \sqcap \neg Pi \sqcap \neg DR & DR \equiv & EC \sqcup DC \\
NTPP \sqsubseteq & \neg TPP \sqcap \neg Pi \sqcap \neg DR & C \equiv & O \sqcup EC & TPP \sqsubseteq & \neg Pi \sqcap \neg DR \\
O \equiv & PO \sqcup P \sqcup Pi & EQ \sqsubseteq & \neg PPi \sqcap \neg DR & Pi \equiv & EQ \sqcup PPi \\
TPPi \sqsubseteq & \neg NTPPi \sqcap \neg DR & P \equiv & EQ \sqcup PP & NTPPi \sqsubseteq & \neg DR \\
PPi \equiv & TPPi \sqcup NTPPi & EC \sqsubseteq & \neg DC & PP \equiv & TPP \sqcup NTPP
\end{array}$$

Figura 7. Representación *exógena* del esqueleto E_1 de RCC.

Concretamente, E es un esqueleto si el propio E junto con los axiomas de nombres únicos, clausura de dominio y completación (los axiomas clásicos de las bases de datos) admite sólo un retículo como modelo.

Formalmente, una teoría es *retículo categórica* (r.c.) si todos sus esqueletos son equivalentes, una vez añadido la citada axiomatización de las bases de datos. Como el retículo es independiente del modelo cuando T es r.c., lo denotaremos por $R(T)$. Éste es el caso de RCC; la distribución que se muestra en la figura 6 es la única posible en los modelos de RCC.

El esqueleto recoge las relaciones conceptuales de la ontología que se desean preservar en la extensión. El diseño del esqueleto de RCC mostrado en la fig. 7 ha sido asistido por MACE4. Nótese que el esqueleto es una descripción de carácter *exógeno*. Es decir, expresa las propiedades de las conexiones espaciales, sólo relaciona entre sí los distintos tipos de conexión. En [8] se define el concepto de extensión retículo-categórica como sigue. Consideremos, para simplificar la notación, que una ontología r.c. O se especifica por un par (T, E) teoría/esqueleto. En adelante, sólo trabajaremos con ontologías r.c. Una ontología $O_2 = (T_1, E_2)$ es una *extensión retículo categórica de* $O_1 = (T_1, E_1)$ (notación: $O_1 \rightarrow_{rc} O_2$) si $R(T_1) \subseteq R(T_2)$ y $R(T_2) \models E_1$. Es decir, extiende el retículo de O_1, respetando las propiedades descritas en el esqueleto E_1.

Consideremos ahora el caso de la fusión de dos ontologías O_1 y O_2 (que, para simplificar, suponemos con lenguajes disjuntos). Supongamos, además, que

disponemos de un conjunto E' de fórmulas que relacionan conceptos de O_1 y O_2. Una ontología r.c. $O = (T, E)$ es una *fusión* de O_1 y O_2 si $O_1 \rightarrow_{rc} O$, $O_2 \rightarrow_{rc} O$ y $R(T) \models E'$, y además $R(T)$ es de tamaño mínimo con esa propiedad. Adicionalmente, si disponemos de una articulación O_a, el diagrama de la fig. 5 (izq.) debe ser conmutativo. El siguiente procedimiento calcula una fusión retículo categórica. Extiende el método de inserción ontológica descrito en [8]:

1. Unir los esqueletos E_1 y E_2 con E', obteniendo un conjunto E_0.
2. Buscar con MACE4 retículos que modelizan a E_0.
3. Si no existen tales retículos, las ontologías son incompatibles con respecto a E_0, y no existe fusión (E_0 es inconsistente). Si existen, pasar a (4).
4. Inspeccionar un retículo de tamaño mínimo. Si alguna de sus relaciones no es aceptada por el usuario, refínese E_0, añadiendo nuevas (in)ecuaciones, para descartar tal retículo. Volver a aplicar MACE4 al refinamiento obtenido. Este paso se repite hasta que un modelo (de tamaño mínimo) sea aceptado por el usuario. De este modo se obtiene un esqueleto S_0.
5. Refinar S_0 hasta que el único retículo de ese tamaño sea el aceptado por el usuario en el paso anterior, obteniendo S.
6. Certificar (con OTTER, si es necesario) que el modelo obtenido es único para S. De esta forma $O = (T_1 \cup T_2 \cup S, S)$ es la fusión deseada.

El paso 5 asegura la retículo-categoricidad de la fusión. Como ya ha comentado comentado, Ocasionalmente es necesaria una etapa adicional, no mecánica, consistente en reinterpretar los términos de las ontologías iniciales [12].

4.1. Fusionando RCC y SIZE

Supongamos que la intención es usar las relaciones de tamaño entre regiones conectadas. La intuición nos dice que se desconoce, a priori, la relación de tamaño entre regiones que no están conectadas. Nótese que esta intencionalidad conlleva una reinterpretación de las relaciones de SIZE. Por ejemplo, $LS(x, y)$ se debe entender como *la región x está conectada a una región de mayor tamaño y*. Esta reinterpretación es compatible con la articulación de la figura 5.

Partimos del esqueleto de SIZE siguiente:

$$E_2 = \begin{cases} \top \equiv LS \sqcup LSi \sqcup SS & LSE \equiv SS \sqcup LS \quad\quad LSi \sqsubseteq LSE \sqcap \neg SS \\ LSi \sqsubseteq \neg SS & \neg SS \equiv LS \sqcup LSi \quad LSEi \equiv SS \sqcup LSi \end{cases}$$

El conjunto de restricciones E_0 estará formado por E' (fig. 4) junto con:

$$\begin{array}{lll} EC \not\sqsubseteq LSE & EC \not\sqsubseteq LSEi & SS \not\sqsubseteq \neg EC \\ DC \not\sqsubseteq LSE & DC \not\sqsubseteq LSEi & SS \not\sqsubseteq \neg DC \end{array}$$

La traza de la ejecución del procedimiento se resume en la siguiente tabla:

Refinamientos	Tamaño mínimo del modelo	número de modelos
sin refinamientos	23	8
$LS \not\sqsubseteq O, LS \not\subseteq O$	24	1
$SS \not\sqsubseteq O$	24	1
$LS \sqcap PO \not\equiv \bot$	27	1
$LSi \sqcap PO \not\equiv \bot$	31	1
$SS \sqcap PO \not\equiv \bot$	31	1

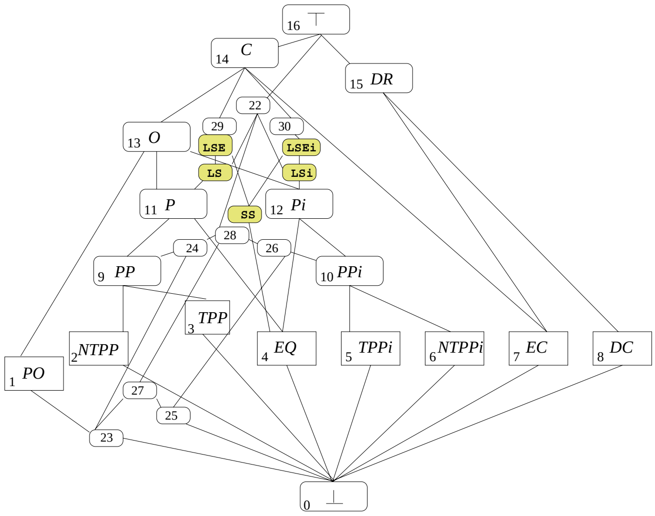


Figura 8. La fusión de RCC y SIZE especializada en conexión

El usuario acepta el modelo final (en la figura 8). Nótese que el usuario puede (y debe) interpretar algunos de los nodos nuevos. Por ejemplo, el nodo 22 representa la relación *tamaños distintos* (i.e. ¬SS), y el nodo 23 representa la relación *x se solapa parcialmente con una región de distinto tamaño y*.

5. Conclusiones y relación con otros trabajos

Hemos presentado un método de fusión centrado en la conceptualización. Existen diferentes aproximaciones al problema, pero en general no usan buscadores de modelos de propósito general para ofrecer fusiones alternativas, como se hace en este artículo.

Por ejemplo, la herramienta Prompt [26] facilita las tareas necesarias para mezclar dos clases en una nueva clase, y localiza las posibles equivalencias. En nuestro método, limitado a la conceptualización, es el SRA el que induce la posible unión, mientras que Prompt pide que se refine el modelo ofrecido por el sistema. Esto puede implicar que el usuario está inconscientemente forzado a mantener relaciones que no son adecuadas. Hcone [23] es, en términos generales, una aproximación parecida. Sin embargo, se basa en el uso de Wordnet para determinar las relaciones entre las dos ontologías, mientras que en nuestro caso, el proceso parte de unas relaciones conocidas y el resto son inducidas por el método para su aprobación por el usuario. ONION [25] también utiliza componentes lingüísticos para calcular la articulación, aunque el proceso es parecido al mostrado en este trabajo. Sin embargo, nuestro método proporciona una ontología con cierto grado de categoricidad. Adicionalmente, al ofertar distintos modelos, se obliga al usuario a refinar el conocimiento para descartar los inadecuados, obteniendo así más información sobre la fusión.

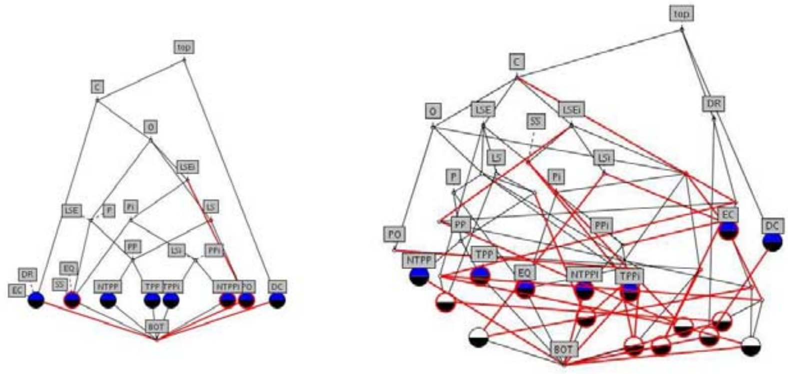


Figura 9. Retículo inicial (izq.) y final (der.) después de la exploración de atributos

En [6] se formaliza la fusión mediante especificaciones algebraicas, siendo la fusión un colímite (en el sentido de teoría de categorías), cuando son compatibles. En nuestro caso, la(s) extensión(es) retículo categórica(s) común(es) obtenida(s) por MACE4 de tamaño mínimo pueden ser consideradas una variante práctica de tal fusión. La compatibilidad viene dada por la existencia de extensiones.

El método FCA-merge [31] es de carácter *extensional*: iguala conceptos con la misma extensión. Es decir, depende de las instancias (la *población* de la ontología). Nuestra propuesta es *intensional* (manejamos los conceptos); puede considerar como una forma de exploración de atributos [17]. Para poder comparar los dos métodos con mayor profundidad, necesitamos una población para la ontología. Hemos utilizado un conjunto de datos sobre relaciones espaciales de provincias, comarcas y regiones del sur de España. Hemos adaptado FCA-merge para evitar el razonamiento lingüístico y sólo hemos utilizado la parte referente a la fusión de los retículos. En la figura 9 se muestra la fusión obtenida en primer lugar y la aceptada después de utilizar la exploración de atributos [17] para refinar el modelo. La conclusión, una vez experimentado con el método, es que la exploración de atributos requiere mucha más reflexión por parte del usuario que el refinamiento exigido en nuestro método. La razón es que este refinamiento está basado en el modelo de tamaño mínimo. Finalmente, una vez acabada la exploración de atributos, el modelo obtenido con FCA-merge contiene un número considerable de nodos no interesantes. Esto es debido a que para el Análisis Formal de Conceptos son interesantes -en teoría- todos los conceptos que se puedan extraer. Nuestro método proporciona una ontología más simple, con menos nodos reduntantes, debido a su tamaño mínimo.

Aplicabilidad y posibles extensiones

Una limitación del método es que el tamaño de las ontologías a fusionar depende de la potencia del buscador de modelos utilizado. Sin embargo, para ontologías espacio-temporales, que son de pequeño tamaño, el método es adecuado con MACE4, pues el usuario puede reconocer la adecuación de la ontología

resultante. Por tanto, es perfectamente aplicable a ontologías temporales como la ya citada de J. Allen y otras similares [21]. Sería interesante diseñar *fusiones contextualizadas* para ontologías de gran tamaño. Una limitación a solventar en el futuro es la extensión del método para trabajar con roles.

Por último, y a la vista del experimento realizado en base al método de FCA-merge, parece adecuado usar la entropía (como en [9]) para seleccionar la fusión adecuada en cada paso del procedimiento en función de los datos.

Agradecimientos

Financiado por el proyecto TIN2004-03884 *Sistemas verificados para el razonamiento en la Web Semántica*, Min. de Ed. y Ciencia (cofinanciado con Fondos FEDER).

Referencias

1. J. F. Allen, Maintaining knowledge about temporal intervals, Communications of ACM, 26(11):832–843 (1983).
2. J. A. Alonso-Jiménez, J. Borrego-Díaz, A. M. Chávez-González and J. D. Navarro Marín, Towards a Practical Argumentative Reasoning with Qualitative Spatial Databases, 6th Int. Conf. on Industrial and Engineering Applications of AI and Expert Systems, Lecture Notes in AI 2718, 789-798, Springer-Verlag 2003.
3. J. A. Alonso-Jiménez, J. Borrego-Díaz and A. M. Chávez-González, Ontology Cleaning by Mereotopological Reasoning Proc. DEXA Workshop on Web Semantics (WebS'04), 2004. IEEE Press, 137-137.
4. J. A. Alonso-Jiménez, J. Borrego-Díaz and A. M. Chávez-González, Foundational challenges in Automated and Ontology Cleaning in the Semantic Web. IEEE Intelligent Systems 21(1):45-52 (2006).
5. G. Antoniou, A. Kehagias, A Note on the Refinement of Ontologies. Int. Journal of Intelligent Systems 15(7): 623-632 (2000).
6. T. Bench-Capon and G. Malcolm, Formalising Ontologies and Their Relations, Proc. of Database and Expert Systems Applications (DEXA 99), LNCS 1677, Springe-Verlag, 1999, pp 250-259.
7. B. Bennett, The Role of Definitions in Construction and Analysis of Formal Ontologies, Proc. of AAAI 2003 Spring Symposium on Logical Formalization of Commonsense Reasoning, 27-35, AAAI Press, 2003.
8. J. Borrego-Díaz and A. M. Chávez-González, Extension of Ontologies Assisted by Automated Reasoning Systems, 10th Int. Conf. on Computer Aided Systems Theory (EUROCAST 2005), LNCS 3643, 247-253, Springer-Verlag, 2005.
9. J. Borrego-Díaz and A. M. Chávez-González, Controlling Ontology Extension by Uncertain Concepts through Cognitive Entropy. ISWC'05 Workshop on Uncertainty Reasoning for the Semantic Web, 56-66 (2005). `http:// sunsite. informatik. rwth-aachen.de/Publications/CEUR-WS/Vol-173/ paper6.pdf`
10. F. Baader, y otros (eds.), The Description Logic Handbook. Theory, Implementation and Applications, Cambridge Univ. Press, 2003.
11. T. Berners-Lee, J. Hendler and O. Lassila, The Semantic Web, Scientific American, May 2001.
12. A. M. Chávez González, Razonamiento Mereotopológico Automatizado para la Depuración de Ontologías, Tesis Doctoral, Univ. de Sevilla (2005).

13. B.L. Clarke, A Calculus of Individuals Based on 'Connection', Notre Dame J. Formal Logic, 22:204-218 (1981).
14. S. Colton, A. Meier, V. Sorge and R. McCasland, Automatic Generation of Clasification Theorems for Finite Algebras, Proc. Int. Joint Conf. on Automated Reasoning (IJCAR 2004), Lecture Notes in AI 3097, 400-414, Springer-Verlag (2004).
15. A. G. Cohn, B. Bennett, J. M. Gooday and N. M. Gotts. Representing and Reasoning with Qualitative Spatial Relations about Regions. chapter 4 in O. Stock (ed.), Spatial and Temporal Reasoning, Kluwer, 1997.
16. M. Cristani and A. G. Cohn, SpaceML: A Mark-up Language for Spatial Knowledge, J. Visual Lang. Comput., 13(1) (2002), 97-116.
17. B. Ganter and R. Wille, Formal Concept Analysis, Mathematical Foundations, Springer, 1999.
18. A. Gerevini and J. Renz, Combining Topological and Size Information for Spatial Reasoning, Artificial Intelligence 137:1-42 (2002).
19. N. M. Gotts, An Axiomatic Approach to Topology for Spatial Information Systems, Report 96.25, School of Computer Studies, Univ. of Leeds, (1996).
20. J. Heflin, Towards the Semantic Web: Knowledge Representation in a Dynamic, Distributed Environment, Ph.D. Thesis, Univ. of Maryland, College Park, 2001.
21. D. Hernández, Qualitative Representation of Spatial Knowledge, Lecture Notes in AI 804, Springer-Verlag, 1994.
22. M. Knauff, R. Rauh and J. Renz, A Cognitive Assessment of Topological Spatial Relations: Results from an Empirical Investigation, Proc. of the 3rd Int. Conf. on Spatial Information Theory (COSIT'97), Lecture Notes in Computer Science 1329, 193-206, Springer-Verlag (1997).
23. K. Kotis and G.A. Vouros, The Hcone approach to Ontology Merging, Proc. of European Semantic Web Symposium (ESWS 2004), Lecture Notes in Computer Science 3053, 137-151, Springer-Verlag, 2004.
24. W. McCune and R. Padmanabhan, Automated Deduction in Equational Logic and Cubic Curves, Lecture Notes in AI 1095, Springer-Verlag, 1996.
25. P. Mitra, G. Wiederhold and M. L. Kersten, A Graph-Oriented Model for Articulation of Ontology Interdependencies, Proc. 7th Int. Conf. Extending Database Tech. (EDBT 2000), LNCS 1777, 86-100, Springer-Verlag (2000).
26. N. F. Noy and M. A. Musen, The PROMPT suite: Interactive tools for ontology merging and mapping. Int. J of Human-Computer Studies, 59(6):983-1024 (2003).
27. J. Renz, Qualitative Spatial Reasoning with Topological Information, Lecture Notes in AI 2293, Springer-Verlag (2002).
28. S. Staab, R. Studer (eds.), Handbook of Ontologies in Information Systems, Springer-Verlag, 2004.
29. J. F. Sowa, Knowledge Representation. Logical, Philosophical and Computational Foundations, Brooks/Cole Pub., 2000.
30. J.G. Stell, Boolean Connection Algebras: A New Approach to the Region-Connection Calculus, Artificial Intelligence 122:111-136 (2000).
31. G. Stumme and Maedche, FCA-Merge: Bottom-Up Merging of Ontologies, Proc. of the 17th Int. Joint Conf. on AI (IJCAI'01), Morgen Kaufmann, 2001.
32. A. N. Whitehead, Process and Reality, MacMillan, New York (1929).
33. F. Wolter and M. Zakharyaschev, Qualitative Spatio-Temporal Representation and Reasoning: a Computational Perspective, in: G. Lakemeyer and B. Nebel (eds.) Exploring AI in the New Millenium, Morgan Kaufmann, 2002, 273-381.

El método del centro de áreas como mecanismo básico de representación y navegación en robótica situada

José R. Álvarez Sánchez, José Mira y Félix de la Paz López

Dpto. de Inteligencia Artificial - UNED. España.
{jras,jmira,delapaz}@dia.uned.es

Resumen Los métodos llamados de potencial y todas sus derivaciones basadas en gradientes, han tenido un uso muy extendido en robótica autónoma, fundamentalmente asociados a estrategias reactivas de navegación. En este artículo presentamos los fundamentos, la formalización y la aplicación de un nuevo método basado en momentos de primer orden llamado "método del centro de áreas". Comentamos también su validez, a nivel individual y combinado con otros métodos, para construir una representación situada del medio.
Palabras clave: robótica autónoma; centro de áreas; representación espacial

1. Introducción

Durante los años que van desde 1985 hasta finales del 2000 son muchas las técnicas que se han usado para intentar mover reactivamente un robot en un medio desconocido, sin necesidad de realizar una representación explícita del mismo. Uno de éstos primeros intentos inspirado en las Matemáticas, fue el llamado espacio de configuraciones (C-Space, ver figura 1.a) [13], en el que se transforma el espacio exterior al robot en líneas formadas por los puntos equidistantes a los obstáculos más cercanos. Esas líneas pueden ser consideradas como posibles trayectorias del robot. También cabe destacar en la misma línea, los trabajos basados en diagramas de Voronoi [6]. Otra técnica, inspirada en éste caso en la Física, es la llamada técnica del potencial repulsivo [4] o campo de fuerzas virtual (VFF). Su fundamento se inspira en las leyes de atracción y repulsión de las cargas eléctricas. Considera al robot y a los obstáculos como cargas de un signo, digamos positivo, y el objetivo al que el robot debe llegar, de signo contrario (negativo). Para mover el robot se plantea un problema simple de composición de fuerzas (suma de vectores), en la que el robot al final se mueve en la dirección, sentido y con la intensidad de la fuerza resultante [5].

Estas y otras técnicas similares han sido utilizadas finalmente como herramientas en arquitecturas más complejas [11,3] que se ocupan de suplir las carencias de éstos métodos, ya que es imposible la navegación eficiente sin inyectar conocimiento externo. Existen otros enfoques para resolver el problema del

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 103-117, 2006.

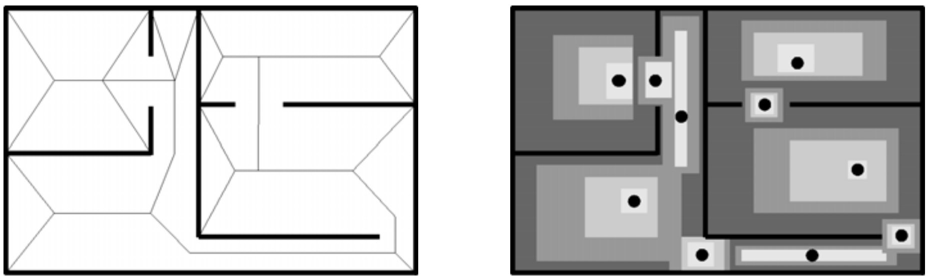
modelado del medio, como por ejemplo los basados en métodos probabilísticos [15,7], pero caen fuera del alcance de nuestra propuesta.

Nuestra propuesta, el método del centro de áreas [2,8,14,10,1,9], puede considerarse como una evolución de esas técnicas. Por un lado es capaz de descubrir esas "rutas seguras" tal y como si fueran trayectorias en C-Space y, por otro lado, se inspira en la Física, en este caso en un momento de primer orden como lo es el centro de masas en Mecánica, añadiendo propiedades de invariancia que el método del potencial no provee, y la posibilidad del uso de la primera derivada (velocidad del centro de áreas respecto al movimiento del robot) para gobernar la navegación.

Nuestro método tampoco es la solución definitiva al problema de la navegación sino que, como veremos al final en el apartado sobre su uso, es una herramienta más que debe ser complementada con conocimiento externo. Nuestro objetivo no es solucionar completamente el problema del modelado del medio por un robot autónomo con éste trabajo, sino dar a conocer una herramienta nueva para el modelado que puede ser utilizada en combinación con otras para llegar a la solución completa, tal y como se muestra en otros trabajos nuestros anteriores [**?**,**?**].

El objetivo de este método es, por tanto, el movimiento guiado por la posición del centro de área de espacio libre detectado. Ese movimiento se realiza dentro de un medio desconocido *a priori* (sin mapa previo y sin marcadores), e incluso contando con la posibilidad de que ocurran ciertos cambios (apertura o cierre de puertas, etc.), o incluso el movimiento de objetos.

Previamente a la descripción del concepto de centro de áreas, realizaremos una descripción de la forma de almacenar, en una representación adecuada, la información captada por los sensores del robot sobre el entorno inmediato local que le rodea, para que se puedan extraer propiedades geométricas aproximadamente invariantes respecto a pequeños cambios de posición locales.



(a) (b)

Figura 1. Tipos de representación del espacio con las técnicas (a) "C-Space" y (b) "centro de áreas".

2. Representación de espacio libre alrededor

Necesitamos una representación del espacio libre alrededor del robot que sea persistente en el tiempo y que acumule la información de los sensores con el movimiento. La representación utilizada está influida y es relativa al contexto del entorno y el robot. Este contexto afectará a las suposiciones y simplificaciones de la representación (y del modelo) del medio. Por ejemplo, aunque hagamos una representación tridimensional completa para un terreno rugoso, si no tenemos sensores dirigidos hacia el suelo, tendremos que suponer un terreno "aproximadamente" plano y los puntos de representación en esa dirección serán estimados por suposiciones incluidas en el modelo del medio (no por lecturas de sensores). O dicho de otra forma, al diseñar un robot para una tarea en un medio, hay que tener en cuenta qué sensores y en qué ubicación se deben de situar para poder representar el medio de forma adecuada para la tarea.

En los ejemplos y nomenclatura que se usarán a continuación, se utiliza principalmente una representación bidimiensional de un espacio tridimensional esencialmente distribuido en un plano, aunque los cálculos y la notación son fácilmente extensibles a una representación tridimensional completa.

2.1. Sensores de rango reales

Son sensores direccionales que devuelven como información, en el instante de lectura, la distancia medida o estimada al objeto más cercano detectable en su rango, según su orientación actual.

La distancia calculada es siempre una medida indirecta a través de la estimación de un modelo que depende del tipo de sensor. Los más habituales de infrarrojos se basan en la intensidad de la luz reflejada por el objeto respecto a la luz emitida (decremento por cuadrado de la distancia), en cambio los típicos de sonar se basan en el tiempo de ida y vuelta del ultrasonido emitido por una membrana (modelos de lóbulos de transmisión, etc.), y los normales de láser se basan en la interferencia entre el haz emitido y el recibido, o bien en modelos de polarización y reflectancia. Podemos considerar diferentes grados de simplificación para estos sensores reales, dependiendo de la complejidad del medio, del tipo de robot y de la tarea que debe realizar. En cualquier caso, esos cálculos y simplificaciones son inherentes al sensor, y en este trabajo consideraremos que forman parte del proceso de detección del sensor y por tanto la información que utilizamos es la distancia correspondiente.

Anchura y sensibilidad del campo de la detección Los sensores pueden detectar objetos en un cono que tiene una anchura limitada y cuyo vértice está centrado en un punto del sensor. La sensibilidad y fiabilidad de la medición (por las estimaciones del modelo interno) pueden variar con la distancia a la que se detecte el objeto. Para simplificar, lo más habitual es considerar la medida de distancia como puntual (y exacta), pero también se podrían tener en cuenta la anchura y sensibilidad del campo de detección e incluir un rango de error en la representación de cada medida.

Posición y orientación respecto al robot Los sensores pueden estar fijos o ser móviles (limitado) respecto al cuerpo del robot. Lo más habitual es que los sensores estén en posiciones fijas del robot y como mucho sean de orientación variable. Como interesa detectar los obstáculos que rodean al robot, los sensores estarán principalmente orientados hacia el exterior del robot. El caso extremo más simple es que todos los sensores estén fijos y orientados radialmente hacia el exterior respecto al centro del robot, aunque la distribución en el perímetro puede no ser uniforme (más sensores delante si el robot no es holonómico) y, además, las distancias respecto al centro pueden ser diferentes dependiendo de la forma exterior del robot.

Llamaremos r a la distancia, respecto a la posición del sensor, de la lectura devuelta por un sensor situado (en el momento de la lectura) en las coordenadas (x_s, y_s) y con una orientación θ_s respecto del robot.

2.2. Transformación a sensores centrados

Necesitamos que la información sobre todos los obstáculos detectados esté representada en el mismo marco de referencia, que consideraremos centrado respecto al robot y en un sistema de coordenadas fijo al cuerpo principal del mismo. Como hemos mencionado antes, la distribución de los sensores en el robot suele ser aproximada a estas características, pero es necesario transformar la información recibida de los sensores sobre los obstáculos a un sistema de referencia único y común. La representación transformada consiste en las coordenadas correspondientes del obstáculo detectado respecto al centro geométrico del robot. Adicionalmente, se mezclan e integran las lecturas de diferentes tipos de sensores de rango respecto a las mismas coordenadas.

Si estamos en un entorno simplificado de tipo 2D$\frac{1}{2}$ (el robot se mueve en un único plano y los obstáculos u objetos son tridimensionales, pero su forma está restringida de manera que el corte por cualquier plano paralelo al suelo sea igual a su proyección sobre éste) las posiciones de los sensores son equivalentes a que todos estén en el mismo plano, dirigidos en paralelo respecto al suelo. En este caso, y dado que estamos representando el mundo desde el punto de vista centrado en el robot, suele ser más conveniente la representación transformada en coordenadas polares. En cambio, en el caso más general de 3 dimensiones sin restricciones sería más conveniente utilizar coordenadas esféricas, o bien algún otro sistema equivalente más cómodo para transformaciones como los cuaterniones.

El valor de una lectura de un sensor real se transforma en coordenadas centradas en el robot si definimos las coordenadas del punto que representa el obstáculo detectado respecto al centro del robot:

$$x_c = x_s + r \cdot \cos\theta_s; \quad y_c = y_s + r \cdot \sin\theta_s \tag{1}$$

Puesto que representamos el espacio alrededor del robot nos interesará el equivalente en polares y orientado respecto al entorno (con el ángulo actual del robot α_R):

$$r_c = \sqrt{x_c^2 + y_c^2}; \quad \hat{\theta}_c = \arctan\left(y_c/x_c\right) + \alpha_R \tag{2}$$

El ángulo α_R de orientación del robot respecto de un sistema de coordenadas fijo en el entorno (exterior al robot) se calcula a partir de la información de sensores de odometría (movimiento de las ruedas) o de sensores inerciales o giroscópicos.

2.3. Interpolación y extrapolación en una distribución uniforme

Necesitamos que la representación del espacio libre alrededor del robot tenga un tamaño fijo (que no crezca indefinidamente) y que, simultáneamente, acumule información por el movimiento de los sensores (respecto al robot) y del robot en el entorno. Para esto debemos ir incorporando, mediante interpolación y extrapolación, los datos de los sensores centrados en una representación con una cantidad fija de puntos en ángulos distribuidos uniformemente alrededor del robot. Podemos denominar *sensores virtuales* a los puntos representados de esta forma, puesto que son equivalentes a unos sensores de rango que estuvieran situados en el centro del robot y dirigidos radialmente y orientados por ángulos equiespaciados.

La representación de espacio libre se debe de ir modificando constantemente en el tiempo para ir guardando y acumulando la información que se recibe en cada instante de las lecturas de los sensores transformadas a las coordenadas comunes centradas en el robot. Lo normal será que una lectura de un sensor corresponda a una orientación que no coincida exactamente con una de las orientaciones fijas de la representación, en cuyo caso la información se debe interpolar y extrapolar junto con los valores guardados en la representación para producir nuevos valores que se almacenaran en las orientaciones más cercanas.

Para acumular una lectura de un sensor real en la representación de las distancias $\{d_i\}$ con distribución en N ángulos uniformes ϕ_i, donde $\phi_i = \frac{2\pi}{N} i$; $i = 0, \ldots, N-1$, se busca el punto correspondiente donde: $\phi_i \leq \hat{\theta}_c < \phi_{i+1}$. Las distancias adyacentes correspondientes d_i y d_{i+1} en la representación se sustituyen por nuevos valores:

$$\hat{d}_i = \mathrm{F}_1\left(d_i, d_{i+1}, r_c, \hat{\theta}_c - \phi_i\right); \quad \hat{d}_{i+1} = \mathrm{F}_2\left(d_{i+1}, d_i, r_c, \phi_{i+1} - \hat{\theta}_c\right) \tag{3}$$

utilizando las funciones de mezcla F_1y F_2 para la interpolación y extrapolación con los valores que ya estaban en ambos casos. El valor de N puede ser (y, de hecho, conviene que lo sea) mayor que la cantidad de sensores reales disponible por el robot, de esta forma se puede cubrir la representación con mayor precisión acumulando información de los sensores desde distintas posiciones (al moverse el robot) y aumentar la resolución de la representación.

2.4. Transformación por movimiento

Cuando el robot se mueve (avance o giro) se debe transformar la representación del espacio libre alrededor del robot a unas nuevas coordenadas que sigan siendo centradas en el robot y con la misma distribución. Es importante hacer notar que la representación está centrada en el robot, pero la orientación de la

representación está fija con respecto al espacio que rodea al robot debido a la inclusión de α_R en la transformación a sensores centrados (apartado 2.2). Es decir, que ahora debemos distinguir en la transformación por movimiento sólo los desplazamientos del centro del robot. Los giros o cambios de orientación no modifican la representación del espacio exterior (sólo cambian los registros internos que llevan cuenta de la orientación del robot respecto a la representación, α_R). La distancia y dirección de desplazamiento del robot respecto al exterior se calcula, análogamente a la orientación, a partir de la información de sensores de odometría (movimiento de las ruedas) o de sensores inerciales o giroscópicos.

La transformación de la representación por traslación del robot se hace, al igual que en la acumulación de nuevos datos de sensores, por interpolación o extrapolación de los anteriores valores en una nueva representación desde la nueva posición del robot. Aunque estos cambios se hagan gradualmente, por pequeños incrementos con el movimiento del robot, es necesario cuidar los casos de oclusión u ocultamiento de unos puntos por otros en la nueva representación, en los cuales el dato que geométricamente queda detrás de otro (según la proyección desde la nueva posición) no contribuye en la nueva transformación.

Si el robot, durante un intervalo de tiempo $\triangle t$, se ha desplazado respecto a un sistema de coordenadas fijo en el entorno en las cantidades

$$\triangle X_R = X_R\left(t\right) - X_R\left(t - \triangle t\right); \quad \triangle X_R = Y_R\left(t\right) - Y_R\left(t - \triangle t\right) \tag{4}$$

debemos transformar la representación centrada para la nueva posición. Para ello se toma cada punto j de la representación anterior $\{(d_j, \phi_j)\}$ y se calculan sus nuevas coordenadas relativas a la posición actual del robot:

$$x_v = d_j \cdot \cos\phi_j - \triangle X_R; \quad y_v = d_j \cdot \sin\phi_j - \triangle Y_R \tag{5}$$

y de forma análoga a como se acumula una nueva lectura de un sensor real, se toman sus equivalentes en polares:

$$r_v = \sqrt{x_v^2 + y_v^2}; \quad \theta_v = \arctan\left(y_v/x_v\right) \tag{6}$$

y se busca el punto correspondiente, donde $\phi_i \leq \theta_v < \phi_{i+1}$, para insertarlo en la nueva representación $\{d_i^*\}$ con las mismas funciones de mezcla, F_1 y F_2, que en el caso del sensor real:

$$\hat{d}_i^* = \mathrm{F}_1\left(d_i^*, d_{i+1}^*, r_v, \theta_v - \phi_i\right); \quad \hat{d}_{i+1}^* = \mathrm{F}_2\left(d_{i+1}^*, d_i^*, r_v, \phi_{i+1} - \theta_v\right) \tag{7}$$

Es importante el orden en el que se realice la inserción. Se debe comenzar con los ángulos ϕ_i más próximos a la dirección correspondiente en la que se ha desplazado el robot. De esta forma se pueden tener en cuenta los puntos que han quedado fuera de la representación a consecuencia del desplazamiento.

Una vez acumulados todos los valores $\{d_j\}$ de la anterior representación, se sustituye por la nueva representación $\{d_i^*\}$ calculada.

3. Concepto de centro de áreas

En Mecánica, la parte de la Física que se encarga de estudiar el movimiento de los cuerpos atendiendo a las causas que lo producen, existe una magnitud denominada centro de masas, que se define como un punto situado a una cierta distancia del eje de movimiento en el que se puede considerar concentrada toda la masa del cuerpo y que permite describir el movimiento de un sólido mediante la descripción del movimiento de ese punto. Para un sistema discreto, la posición del centro de masas es la suma de las coordenadas de cada partícula que compone el sistema ponderadas por su masa relativa a la masa total del sistema. Puesto que se trata de una suma ponderada, se pueden calcular los centros de masas de grupos de partículas y después aplicar el mismo cálculo con los grupos (con la masa correspondiente) y el resultado del centro de masas total es el mismo.

Este concepto de centro de masas se puede aplicar al modelado del medio por un robot, pero sustituyendo la masa por el área accesible detectada. Desde el punto de vista de una representación del espacio por puntos del entorno distribuidos en N ángulos (de la sección 2), se puede interpretar el espacio alrededor como un polígono (las distancias al centro del robot marcan los vértices), que se subdivide fácilmente en triángulos (uniendo cada par de vértices adyacentes con el origen de coordenadas centrado en el robot). Para calcular el centro de áreas es más cómodo calcular primero los centros de áreas de cada uno de esos triángulos y después el centro de área total. El centro de áreas de un triángulo está en su baricentro o centroide, que está a 2/3 del promedio (ó 1/3 de la suma vectorial) de dos lados medido desde el vértice que los une. Por tanto para cada triángulo, entre ϕ_i y ϕ_{i+1} (donde $\phi_i - \phi_{i+1} = \frac{2\pi}{N}$),

$$x_i = \tfrac{1}{3}\left(d_i \cos\phi_i + d_{i+1}\cos\phi_{i+1}\right); \quad y_i = \tfrac{1}{3}\left(d_i \sin\phi_i + d_{i+1}\sin\phi_{i+1}\right) \tag{8}$$

son las coordenadas (x_i, y_i) de su baricentro y su área es $s_i = \frac{1}{2} d_i d_{i+1} \sin\frac{2\pi}{N}$.

Las coordenadas del centro de área total son simplemente las sumas de las coordenadas de los triángulos ponderadas por sus áreas relativas, es decir,

$$x_A = \frac{\sum_i x_i s_i}{\sum_i s_i}; \quad y_A = \frac{\sum_i y_i s_i}{\sum_i s_i} \tag{9}$$

que son las coordenadas relativas respecto del centro del robot pero con orientación absoluta respecto del exterior.

En el caso más general en 3D, la representación de distancias alrededor serían interpretables como los vértices de un poliedro del cual podríamos calcular el centro de volúmenes de forma similar descomponiéndolo en pirámides partiendo desde el origen de coordenadas (el robot).

Para que la información del centro de áreas (o de volúmenes) sea útil, es necesario tener la máxima información posible distribuida alrededor del robot en todas las direcciones. Esto se puede conseguir utilizando muchos sensores por todo el perímetro exterior del robot, o bien moviendo (girando) el robot completo o una plataforma de sensores (como la torreta giratoria del robot Nomad-200 y similares).

3.1. Movimiento del centro de áreas

El baricentro geométrico de una superficie acotada exacta dada (una porción del mapa en el dominio de un observador externo) es fijo y único. La representación de un entorno obtenida a partir de sensores de rango, tal como hemos descrito en la sección 2, puede ser ligeramente diferente (aunque no demasiado) dependiendo del punto donde está situado el robot (e incluso de la trayectoria recorrida para llegar a él). Es usual que las superficies representadas no estén totalmente acotadas, sólo están limitadas por el alcance de los sensores y de las oclusiones de ciertas partes, o por cambios debidos a objetos móviles. Por tanto, al obtener la representación de una zona desde distintos puntos dentro de la misma, puede haber variaciones que afecten a la posición obtenida para el centro de áreas calculado usando cada representación (en el dominio propio).

Hay que tener en cuenta que también puede haber pequeñas diferencias en varias representaciones hechas desde un mismo punto que dependen de la trayectoria de llegada a ese punto y de la calidad o distribución de los sensores de rango reales (por ejemplo, en robots que no tienen sensores distribuidos alrededor de todo su perímetro o que no tienen una plataforma giratoria para mover los sensores). A pesar de estas pequeñas diferencias en la representación obtenida desde un mismo punto, la posición calculada para el centro de áreas es bastante robusta y varía muy poco si el entorno no ha cambiado realmente (solo la trayectoria de llegada al punto desde donde se calcula). Por tanto, podemos hacer una correspondencia de cada punto o posición actual del robot (en el cual tenemos una representación del entorno) con la posición del centro de áreas calculado correspondiente, considerando ambas en el sistema de coordenadas referido al entorno.

Cuando el entorno circundante es muy cerrado (los obstáculos están en su mayoría dentro del alcance de los sensores) se observa que, usando las representaciones transformadas obtenidas desde diferentes puntos dentro de ese mismo entorno, la posición obtenida para el centro de áreas es aproximadamente la misma. Es decir, que la posición calculada para el centro de áreas es aproximadamente invariante cuando se calcula a partir de representaciones adecuadas del entorno obtenidas desde diferentes puntos "cercanos" del mismo. Para ese tipo de entornos locales en los cuales el centro de áreas es casi el mismo, independientemente de la posición del robot desde la que se obtiene la representación usada para calcularlo, decimos que el centro de áreas es estable.

Para poder obtener una medida de esa característica del entorno con respecto del centro de áreas (estabilidad o invariancia), utilizamos la idea de ver cómo varía la posición del centro de áreas en relación a la variación de posición del robot desde la cual lo calculamos. No disponemos de los datos correspondientes a todos los pares de posición robot y posición del centro de áreas calculado desde ella (ni siquiera de un amplio subconjunto). Dado que el robot se mueve siempre en una trayectoria y la actualización de la representación respecto a su movimiento se hace por pasos discretos, podemos utilizar esos pasos o incrementos para calcular por diferencias el movimiento relativo de la posición del centro de áreas correspondiente. Necesitamos comparar las posiciones (absolutas

respecto al entorno), tanto del robot como del centro de áreas correspondiente a la representación desde ese punto, en los instantes t y $t - \triangle t$ (actual y anterior).

La variación de posición del robot en un paso, usada en la ec. 4, tiene su correspondiente para el cambio de posición del centro de áreas respecto las coordenadas externas

$$\triangle X_{CA} = X_{CA}(t) - X_{CA}(t - \triangle t)\,; \quad \triangle Y_{CA} = Y_{CA}(t) - Y_{CA}(t - \triangle t) \qquad (10)$$

que a su vez dependen de las posiciones relativas de los centros de áreas respecto al centro del robot (de las ec. 9)

$$\begin{array}{ll} X_{CA}(t) = X_R(t) + x_A(t)\,; & X_{CA}(t - \triangle t) = X_R(t - \triangle t) + x_A(t - \triangle t) \\ Y_{CA}(t) = Y_R(t) + y_A(t)\,; & Y_{CA}(t - \triangle t) = Y_R(t - \triangle t) + y_A(t - \triangle t) \end{array} \qquad (11)$$

La forma en la que varía la posición del centro de áreas, calculado al variar la posición del robot, depende de la dirección del movimiento que hace el robot con respecto a la línea que lo une con el centro de áreas. Según cómo sea el movimiento, esencialmente longitudinal o transversal a esa linea hacia el centro de áreas, podremos distinguir diferentes tipos de entornos locales donde está situado el robot.

Descomponemos el movimiento del robot y el correspondiente del centro de áreas, según las proyecciones perpendiculares y longitudinales respecto a la línea que une ambos en el punto de partida (en $t - \triangle t$), y definimos

$$\begin{array}{l} \triangle_R^{\parallel} = \triangle X_R \cdot x_A(t - \triangle t) + \triangle Y_R \cdot y_A(t - \triangle t) \\ \triangle_R^{\perp} = \triangle Y_R \cdot x_A(t - \triangle t) - \triangle X_R \cdot y_A(t - \triangle t) \end{array} \qquad (12)$$

$$\begin{array}{l} \triangle_{CA}^{\parallel} = \triangle X_{CA} \cdot x_A(t - \triangle t) + \triangle Y_{CA} \cdot y_A(t - \triangle t) \\ \triangle_{CA}^{\perp} = \triangle Y_{CA} \cdot x_A(t - \triangle t) - \triangle X_{CA} \cdot y_A(t - \triangle t) \end{array} \qquad (13)$$

donde $\triangle_R^{\parallel}$ y $\triangle_R^{\perp}$ son las componentes longitudinal y transversal respectivamente del movimiento del robot, y $\triangle_{CA}^{\parallel}$ y $\triangle_{CA}^{\perp}$ son las mismas pero del movimiento del centro de áreas calculado. Para tener las proyecciones normalizadas habría que dividir todas por $D_{CA} = \sqrt{(x_A(t - \triangle t))^2 + (y_A(t - \triangle t))^2}$, la distancia del robot al centro de áreas en $t - \triangle t$, pero como luego se usan unas en relación con otras, se puede simplificar la definición. Por tanto, estas definiciones son válidas sólo si el robot **no** está justo sobre el centro de áreas.

3.2. Invariancia o estabilidad del centro de áreas

Cuando la posición del robot no coincide justo en la misma posición que el centro de áreas calculado correspondiente, podemos observar esencialmente tres tipos de entorno que dependen de la distribución de los obstáculos detectados alrededor del robot. En la figura 2 se muestran unos ejemplos de los tipos de patrones de movimiento relativo entre el robot y el centro de áreas calculado, para los tres tipos de entorno que denominamos estable, tránsito e inestable.

En la figura se han representado mediante números encerrados en cuadrados las posiciones del robot y mediante los mismos números pero encerrados en círculos las posiciones correspondientes del centro de área.

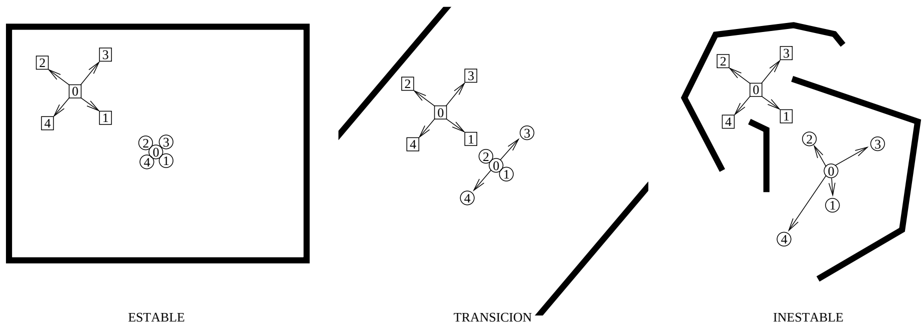


Figura 2. Patrones de correspondencia del centro de áreas (números en círculos) respecto a las posiciones del robot (números en cuadrados) desde donde se calculan en entornos locales de tipo estable, transición e inestable, usados en el cuadro 1.

Podemos observar que dependiendo de que los movimientos del robot sean longitudinales o transversales respecto de la posición inicial del robot en cada tipo de entorno, los movimientos longitudinales o transversales del centro de áreas son diferentes. Hemos resumido las diferentes posibilidades en el cuadro 1. En ese cuadro se han indicado algunos desplazamientos del centro de áreas próximos a cero en algunos casos, pero en la práctica para clasificar la zona se utilizan umbrales sobre la relación del movimiento del centro de áreas divido por el movimiento (longitudinal y transversal) que se haya realizado con el robot (velocidad relativa del CA respecto al robot). En el cuadro 1 también hemos añadido una línea más donde se indican los criterios correspondientes si la posición de partida del robot coincide con el centro de áreas calculado.

4. Uso del método del centro de áreas

La información sobre el espacio libre que rodea al robot siempre es local e instantánea y por tanto sólo da una visión centrada en la posición actual del robot. Si se quiere guardar más información relativa a los puntos anteriormente visitados por el robot debe hacerse en una estructura de otro nivel superior. Como sería ineficiente ir guardando todas las sucesivas representaciones locales desde los puntos por los que ha pasado el robot, es preferible guardar únicamente algunas características representativas (más o menos invariantes) de ciertos puntos relevantes (por alguna propiedad interesante de esas características) y las relaciones topológicas (conectividad de paso para el robot) entre esos puntos.

Cuadro 1. Clasificación de zonas comparando las componentes longitudinal o transversal (respecto a la línea robot-CA) del movimiento del centro de áreas en relación al movimiento del robot. Los nombres de puntos de movimiento ($\boxed{0} \rightarrow \boxed{1}$, etc.) siguen los patrones mostrados en la figura 2.

Clasificación zona (cuando $D_{CA} > 0$)				Estable		Transición		Inestable	
Movimiento robot relativo al CA				$\triangle^{\parallel}_{CA}$	$\triangle^{\perp}_{CA}$	$\triangle^{\parallel}_{CA}$	$\triangle^{\perp}_{CA}$	$\triangle^{\parallel}_{CA}$	$\triangle^{\perp}_{CA}$
longitudinal	$\boxed{0} \rightarrow \boxed{1}$	se acerca	$\triangle^{\parallel}_{R} > 0$	≈ 0	≈ 0	≈ 0	≈ 0	$\sim \triangle^{\parallel}_{R}$	$\neq 0$
$\triangle^{\perp}_{R} \approx 0$	$\boxed{0} \rightarrow \boxed{2}$	se aleja	$\triangle^{\parallel}_{R} < 0$	≈ 0	≈ 0	≈ 0	≈ 0	$\sim \triangle^{\parallel}_{R}$	$\neq 0$
transversal	$\boxed{0} \rightarrow \boxed{3}$	hacia la izq.	$\triangle^{\perp}_{R} > 0$	≈ 0	≈ 0	≈ 0	$\approx \triangle^{\perp}_{R}$	$\neq 0$	$\sim \triangle^{\perp}_{R}$
$\triangle^{\parallel}_{R} \approx 0$	$\boxed{0} \rightarrow \boxed{4}$	hacia la dcha.	$\triangle^{\perp}_{R} < 0$	≈ 0	≈ 0	≈ 0	$\approx \triangle^{\perp}_{R}$	$\neq 0$	$\sim \triangle^{\perp}_{R}$
Si el robot está en CA ($D_{CA} \approx 0$) y se mueve				No cambia		Dep. direcc.		Se mueve	

Por ejemplo, se puede utilizar una representación transformada respecto al centro de áreas, que se calcula de igual forma que la transformación por movimiento del apartado 2.4, pero sustituyendo en las ecuaciones 5 la posición relativa del centro de áreas calculado en la ec. 9, (x_A, y_A), en lugar de la cantidad que se ha desplazado el robot $(\triangle x_R, \triangle y_R)$. Esta nueva representación transformada, con origen en el centro de áreas, tiene propiedades útiles cuando se desea comparar un entorno almacenado en el control de alto nivel, pues es aproximadamente invariante en una amplia zona donde el centro de áreas calculado es el mismo.

Otra propiedad del centro de áreas útil en el control de alto nivel es que representa siempre un punto “seguro” para el robot, pues está libre de obstáculos (salvo casos anómalos, fácilmente detectables y distinguibles, en los que la representación no es adecuada). Entonces se puede ir navegando moviendo el robot de un centro de áreas estable hacia otro.

El método de centro de áreas (CA), tal y como hemos comentado, es una herramienta que puede ser utilizada por otras instancias de alto nivel para la navegación efectiva de robots en medios desconocidos y cambiantes. Éste método tiene ventajas significativas frente a otros métodos similares como el método del campo de fuerzas virtual (VFF) [4], pues proporciona herramientas para poder salvar los denominados “mínimos locales”. El método VFF fue concebido en principio como un método global de solución a la navegación de un robot entre dos puntos pues no sólo gobernaba el movimiento del robot mediante la fuerza resultante, sino que permitía al robot llegar a una meta preestablecida. Sin embargo, como nos enseña la Física elemental, existen muchos puntos en ese espacio donde el robot se mueve donde la resultante de las fuerzas de atracción y repulsión es nula, lo que hace que el robot se pare o realice ciclos periódicos de movimiento que le impiden, en definitiva, llegar a su meta. El método CA en sí, sin otros métodos de nivel superior que lo usen, no soluciona éste problema pues cada centro es propiamente un mínimo local, pero sí proporciona, mediante las

propiedades de estabilidad y el cálculo de la velocidad relativa, las herramientas necesarias para que otros métodos de alto nivel gobiernen el movimiento de manera fiable, segura y efectiva. En la figura 3 podemos ver una ilustración del problema de los mínimos locales en ambos métodos.

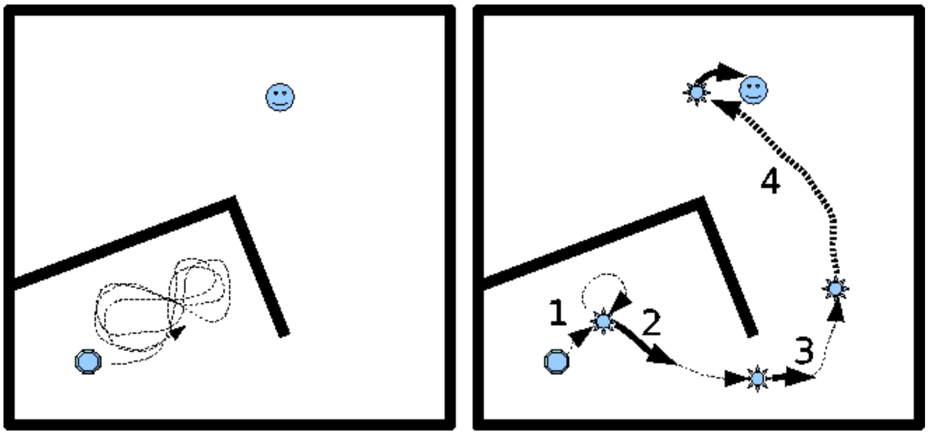


Figura 3. Comparación del problema de mínimos locales entre el método VFF (izq.) y el del centro de áreas (dcha.).

Para el caso del método VFF (figura 3, izquierda), el robot, representado por un octógono, queda confinado siempre en la misma zona, recorriendo un circuito cíclico, por lo que, a menos que se utilice algún otro método adicional, nunca alcanzara la meta, representada por un icono sonriente.

En el caso del método CA (figura 3, derecha):

1. El robot navega reactivamente al centro de áreas más cercano a su posición inicial, representado por una estrella. Esta navegación reactiva, al igual que en caso del VFF, la representamos por una flecha fina discontinua.
2. Después, utilizando las propiedades de estabilidad y el calculo de la velocidad relativa, podemos establecer un método de alto nivel para sacarlo de ese centro y que representamos por una flecha gruesa continua. Una vez fuera del centro, navega reactivamente hacia el siguiente centro de áreas.
3. De nuevo, utilizamos un método de alto nivel para sacarlo del centro en la dirección adecuada.
4. La línea gruesa discontinua representa a los centros de área inestables que se suceden en esa zona. Finalmente el robot alcanza la meta deseada.

5. Conclusiones

Hemos presentado en este trabajo el método del centro de áreas y su potencial validez, de forma individual y combinada con otros métodos, para construir

Cuadro 2. Comparación de características de los métodos de potencial (VFF) y del centro de áreas (CA).

	VFF	CA
Grado de autonomía	Alto, no necesita técnicas de alto nivel	Bajo
Trayectorias en el medio	Óptimas si no hay mínimos locales	No óptimas, pero razonables
Mínimos locales	son un problema	CA=mínimo local (los usa para moverse)
Complejidad del medio	Baja, problema con las simetrías	Alta
Motivación del movimiento	Repulsión de obstáculos	Búsqueda de espacio libre
Integración con otros métodos	Buena, con cualquiera de alto nivel	Buena, con cualquiera de alto nivel
Representación	Sin representación explícita	Mapas topológicos y/o grafos
Paradigmas	Reactivo Híbrido	Híbrido

una representación situada del medio. A lo largo del desarrollo de este método hemos reflexionado sobre la idea de "*complejidad del medio*" y sus exigencias en la complejidad de un sistema (un "agente") que tenga que moverse en ese medio. En términos evolutivos, este acoplo estructural entre un sistema y su medio ha modulado la deriva de ambos, de forma que cada "complejidad" de un medio ha obligado a los sistemas, que pretendían sobrevivir en el mismo medio, a adaptarse desarrollando un lenguaje interno de representación, que es el resultado de la *abstracción* de las propiedades de ese medio que fueran relevantes para su supervivencia. Estas propiedades son las que le han garantizado la ventaja evolutiva de *reaccionar* de forma eficiente en tiempo real *reduciendo la dimensionalidad* del espacio y refinando los mecanismos de atención selectiva. El ejemplo de la rana *pipiens* [12] es paradigmático de esta adaptación.

Si sustituimos ahora a la evolución por el intento de diseño de un robot capaz de navegar en un entorno concreto, estamos obligados a inyectarle el conocimiento sintético equivalente al de la evolución. Es decir, somos *nosotros* los que necesitamos *abstraer* las características relevantes de ese medio para la tarea concreta a la que se va a dedicar el robot, y dotarle del mecanismo o conjunto de mecanismos de representación interna adecuados para su acoplo estructural a ese medio.

Así, la idea subyacente al método del centro de áreas que hemos expuesto en este trabajo, nace del siguiente razonamiento. El robot sólo podrá moverse por zonas de espacio abierto próximas a su posición en cada momento. Por consiguiente, el mecanismo básico en su representación del medio deberá de proporcionarle esencialmente información sobre esas zonas de espacio abierto. Si un observador externo con una visión global del medio tuviera que darle

"*instrucciones verbales*" a un supuesto robot "ciego", por *gestalt* buscaría el centro de la zona más despejada alrededor del robot y le daría "instrucciones situadas", suponiendo que el observador está sobre el robot y que este dispone de un sistema de referencia local que le permite entender "hacia adelante", "despacio", "un poco a tu derecha", "para", "retrocede", "gira algo a tu izquierda", "¡ahí, ahí!", "avanza recto", etc. El observador podría afirmar que "es el medio el que mueve al robot". Las áreas libres son el único espacio de "opciones válidas" para el desplazamiento. Sobre este espacio hay que superponer las trayectorias de acuerdo con otros objetivos adicionales a la mera navegación ("ir a (x,y)", "coger A", ...). El resto del espacio no existe, en sentido estricto. Al robot no le interesan los obstáculos, sino su complementario.

Agradecimientos

Agradecemos la financiación de este trabajo proporcionada por el Ministerio de Ciencia y Tecnología con el proyecto AVISA (Proyecto Coordinado de I+D MCyT, ref. TIN2004-07661-C02-01).

Referencias

1. José R. Álvarez, Félix de la Paz, y José Mira. On Virtual Sensory Coding: An Analytical Model of the Endogenous Representation. En José Mira y Juan V. Sánchez-Andrés, directores, *Engineering Applications of Bio-Inspired Artificial Neural Networks*, volúmen 1607 de *Lecture Notes in Computer Science*, páginas 526–539. International Work-Conference on Artificial and Natural Neural Networks, IWANN'99, Springer-Verlag, junio 1999.
2. José R. Álvarez-Sánchez, Félix de la Paz, y José Mira. A Robotics Inspired Method of Modeling Accesible Open Space to Help Blind People in th Orientation and Traveling Tasks. En José Mira y Jose R. Álvarez, directores, *Mechanisms, Symbols, and Models Underlying Cognition*, volúmen 3561 de *Lecture Notes in Computer Sciences*, páginas 405–415. International Work-conference on the Interplay between Natural and Artificial Computation, IWINAC 2005, Springer-Verlag, junio 2005.
3. Ronald C. Arkin. *Behavior Based Robotics.* M.I.T. Press, 1998.
4. Johann Borenstein. Real Time Obstacle Avoidance for Fast Mobile Robot. *IEEE Transactions on System, Man and Cybernetics*, 19(5):1179–1187, septiembre 1989.
5. Johann Borenstein y Yoram Koren. Histogramic In-Motion Mapping for Mobile Robot Obstacle Avoidance. *IEEE Transactions on Robotics and Automation*, 7(4):535–539, agosto 1991.
6. J. Canny y B. Donald. Simplified Voronoi Diagrams. *Discrete and Computational Geometry*, 3(3):219–236, 1988.
7. Howie Choset, Kevin M. Lynch, Seth Hutchinson, George Kantor, Wolfram Burgard, Lydia E. Kavraki, y Sebastian Thrun. *Principles of robot motion.* MIT press, 2004.
8. F. de la Paz, J. R. Álvarez, y J. Mira. An Analytical Method for Decomposing the External Environment Representation Task for a Robot with Restricted Sensory Information. En Changjiu Zhou, Dario Maravall, y Da Ruan, directores,

Autonomous Robotic Systems: Soft Computing and Hard Computing Methodologies and Applications., volúmen 116 de *Studies in Fuzziness and Soft Computing*, páginas 189–215. Physica-Verlag, 2003.

9. Felix de la Paz López. *Una arquitectura que integra el modelado endógeno del medio y la navegación para un robot genérico de ruedas.* tesis doctoral, E. T. S. I. Informática - UNED, 3 abril 2003.
10. Félix de la Paz López y José Ramón Álvarez Sánchez. Topological Maps for Robot's Navigation: A Conceptual Approach. En José Mira y Alberto Prieto, directores, *Bio-Inspired Applications of Connectionism*, volúmen 2085 de *Lecture Notes in Computer Science*, páginas 459–467. International Work-Conference on Artificial and Natural Neural Networks, IWANN 2001, Springer-Verlag, junio 2001.
11. Benjamin Kuipers y Yung-Tai Byun. A Robot Exploration and Mapping Strategy Based on a Semantic Hierarchy of Spatial Representations. *Journal of Robotics and Autonomous Systems*, (8):47–63, 1991.
12. J.Y. Lettvin, H.R. Maturana, W.S. McCulloch, y W.H. Pitts. What the Frog's Eye Tells the Frog's Brain. En *Proceedings of the IRE*, volúmen 47, páginas 1940–1951. 1959.
13. Tomás Lozano-Perez. Spatial Planning: A Configuration Space Approach. *IEEE Transactions on Computers*, 32(2):108–120, 1983.
14. Jose Mira, José Ramón Álvarez Sánchez, y Félix de la Paz López. The Knowledge Engineering Approach to Autonomous Robotics. En José Mira y José R. Álvarez, directores, *Artificial Neural Nets Problem Solving Methods*, volúmen 2687 de *Lecture Notes on Computer Science*, páginas 161–168. International Work-Conference on Artificial and Natural Neural Networks, IWANN 2003, Springer-Verlag, junio 2003.
15. S. Thrun. Robotic Mapping: A Survey. En G. Lakemeyer y B. Nebel, directores, *Exploring Artificial Intelligence in the New Millenium.* Morgan Kaufmann, 2002.

Localización de fuentes del conocimiento en el proceso del mantenimiento del software

Juan P. Soto[1], Oscar M. Rodríguez[2], Aurora Vizcaíno[1], Mario Piattini[1] y Ana I. Martínez-García[2]

[1] Grupo Alarcos, Departamento de Tecnologías y Sistemas de Información
Ciudad Real (España)
jpsoto@proyectos.inf-cr.uclm.es
{Aurora.Vizcaino,Mario.Piattini}@uclm.es
[2] CICESE, Departamento de Ciencias Computacionales
Ensenada, México
{orodrigu,martinea}@cicese.mx

Resumen. En toda organización es fundamental que las fuentes de información y conocimiento sean localizables en el momento en que son requeridas. Particularmente, en el mantenimiento del software (MS) sería conveniente poder localizarlas fácilmente debido a la cantidad de conocimiento que se necesita durante la ejecución de este proceso. Este artículo presenta una ontología de fuentes de conocimiento que permite representar y localizar el conocimiento requerido por los ingenieros del mantenimiento. Agilizando, de esta forma, el proceso de MS, ya que se da soporte en las tareas diarias de los encargados del mantenimiento al facilitarles el acceso al conocimiento necesario para el desarrollo de sus actividades.

Palabras clave: Ontologías, gestión del conocimiento, mantenimiento del software

1 Introducción

La Gestión del Conocimiento (GC) es una disciplina que promete sacar provecho del capital intelectual de las organizaciones [15], al proporcionar métodos que simplifican procesos como la "compartición", distribución, creación, almacenamiento, organización y entendimiento del conocimiento de una compañía [1, 7]. En la ingeniería de software, y en particular en la etapa de mantenimiento, las técnicas de GC han causado gran expectación debido a la cantidad de conocimiento que se necesita para la ejecución de este proceso. La documentación relacionada con un sistema software al que hay que darle mantenimiento, comúnmente es escasa u obsoleta y casi nunca es actualizada conforme el sistema evoluciona. Algunos estudios indican que del 40% al 60% del esfuerzo por mantener el software está dedicado a entender el sistema [11]. Con frecuencia, las compañías cuentan con documentos o personas con información o conocimiento necesarios para ayudar en sus actividades a los ingenieros de mantenimiento, sin embargo éstos ignoran su existencia o localización.

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 118-123, 2006.

Para resolver algunos de estos problemas se ha planteado usar técnicas de GC. Sin embargo, esto conlleva otros problemas, como evitar sobrecargar a los empleados con nuevas tareas tales como la captura de información en un sistema de GC. Otro problema importante es definir el tipo de información y conocimiento que la empresa posee y dónde está localizado, ya que la principal barrera para compartir el conocimiento es la "ignorancia" de su existencia [17]. Además, por lo general las organizaciones no saben como localizar a las personas expertas que poseen el conocimiento necesario para resolver determinado problema [10]. Para poder paliar este problema hemos definido dos ontologías, la primera de ellas nos permite definir las fuentes de conocimiento con que cuenta una compañía y la otra nos permite especificar la información y conocimiento que puede ser obtenido de cada una de las fuentes de conocimiento. Para probar nuestra propuesta hemos desarrollado ambas ontologías en el dominio del Mantenimiento de Software (MS).

El resto de este artículo está organizado de la siguiente forma: en la sección 2 se describen los problemas que surgen en el mantenimiento y las ontologías vinculadas a éste. La sección 3 ilustra la ontología de fuentes de información para el MS. Finalmente se presentan las conclusiones.

2 Mantenimiento del software y ontologías

El MS ha sido definido como "la modificación de un producto software después de haber sido entregado a los usuarios o clientes con el fin de corregir defectos, mejorar el rendimiento u otros atributos, o adaptarlo a un cambio en el entorno" [9]. El proceso de MS es el que más trabajo e información requiere durante el ciclo de vida del software. Existen varios motivos que complican este proceso, tales como: la falta de documentación asociada al producto a mantener, hacer mantenimiento de manera "ad hoc", es decir, en un estilo libre establecido por el propio programador, la falta de metodologías que den soporte a este proceso, etc.

Un sistema de GC podría ayudar a evitar algunos de los problemas comentados previamente. Por ejemplo, si las organizaciones almacenaran su información y conocimiento en dicho sistema, estas podrían retener el capital intelectual y buenas prácticas de sus empleados. Además, si un trabajador deja la empresa, su experiencia y conocimiento se queda en la organización. De esta forma se evitará la repetición de errores e incrementaría la productividad y probabilidad de éxito [8].

Antes de construir un sistema de gestión del conocimiento para el MS, es de vital importancia modelar, estructurar y generalizar la información que se genera y consulta durante el proceso de MS. Para ello, hemos utilizado ontologías, ya que permiten hacer una especificación explícita de una conceptualización [3]. Las ontologías pueden ser utilizadas para compartir el conocimiento de la organización, así como para promover la interoperabilidad entre sistemas. Diferentes autores han planteado ontologías relacionadas con el MS. En la ontología de Kitchenham et al. [4] se describen los principales aspectos que deben ser tenidos en cuenta en la realización de estudios empíricos en el MS. Ruiz et al. [14] presentan una ontología orientada a la gestión de proyectos de MS, la cual busca representar los aspectos estáticos y dinámicos del proceso de MS desde un punto de vista de los procesos de negocio. Por su parte, Días

et al. [2] desarrollaron una ontología para describir el conocimiento utilizado en el MS. Todas estas ontologías categorizan algunos tipos de fuentes de conocimiento. Sin embargo, los autores no explican dónde localizar estas fuentes y cómo consultarlas, siendo estos los aspectos más importantes para la GC.

3 Ontología de las fuentes de conocimiento en el mantenimiento del software

Las fuentes de información o conocimiento consultadas durante el MS pueden ser muy diversas. Cada organización es muy distinta y tiene diferentes modos de manejar su información. Sin embargo, es posible hacer algunas generalizaciones al respecto de las fuentes de conocimiento que consultan los encargados del MS [5, 7, 17].

Días et al. [2] agrupan los tipos de documentos utilizados en el MS, en tres categorías: 1) los que son parte del producto o sistema a mantener, tales como las especificaciones de requerimientos, diseño (lógico o físico), y producto; 2) documentos del proceso, como planes de pruebas, configuración, aseguramiento de la calidad, y desarrollo de software; y 3) documentos de soporte, como manuales de usuario, hardware, operación, y mantenimiento.

Seaman [16] presenta un estudio enfocado en identificar las estrategias que los encargados del MS emplean para obtener información, además mencionan las fuentes de información utilizadas en este proceso.

Lethbridge et al. [6] presentan un estudio sobre las estrategias de uso de documentación de los ingenieros de software. Los tipos de documentos mencionados en este estudio son: documentación de pruebas o calidad, diseño de bajo nivel, requerimientos, arquitectura, diseño detallado, y especificaciones. Entre los resultados interesantes de este estudio se encuentra que aun cuando la documentación con frecuencia no está actualizada, ésta sigue siendo útil en muchos casos. Finalmente, Koskinen et al. [5] estudiaron qué información es necesaria para comprender un programa. Los autores clasifican las fuentes de información en tres categorías principales: 1) personas, 2) herramientas de soporte, y 3) fuentes de información obtenida fuera del entorno de mantenimiento.

Teniendo en cuenta los trabajos anteriores hemos definido una taxonomía para clasificar las fuentes de conocimiento más importantes que un mantenedor suele consultar. Esta taxonomía se descompone en 4 tipos de fuentes: 1) documentación, 2) sistema o producto, 3) personas, y 4) herramientas de soporte (ver Figura 1).

La **documentación** puede dividirse en:

- *Documentación del sistema*, todos aquellos documentos que describen los productos que son mantenidos.
- *Documentación técnica* está compuesta de todos aquellos documentos relacionados con los lenguajes de programación, herramientas de desarrollo, etc. utilizado por los mantenedores para llevar a cabo su trabajo. Por ejemplo, manuales, tutoriales, libros, etc., acerca de una herramienta de desarrollo o lenguaje en específico.
- *Documentación de usuario*, los documentos creados para los usuarios o clientes del sistema mantenido, por ejemplo: manuales de instalación o configuración.

– *Documentación organizacional*, los documentos relacionados con la vida de la organización, tales como su estructura organizacional, sus normas y políticas, descripción de procesos, etc.
– *Documentación del proceso de mantenimiento*, documentación relacionada con el proceso de MS, tales como peticiones de mantenimiento, planes de prueba y reportes, de entrega, de gestión de la configuración, aseguramiento de calidad, etc.
– *Otros documentos*, esta categoría es utilizada para clasificar todos los documentos que no han sido considerados en las otras categorías.

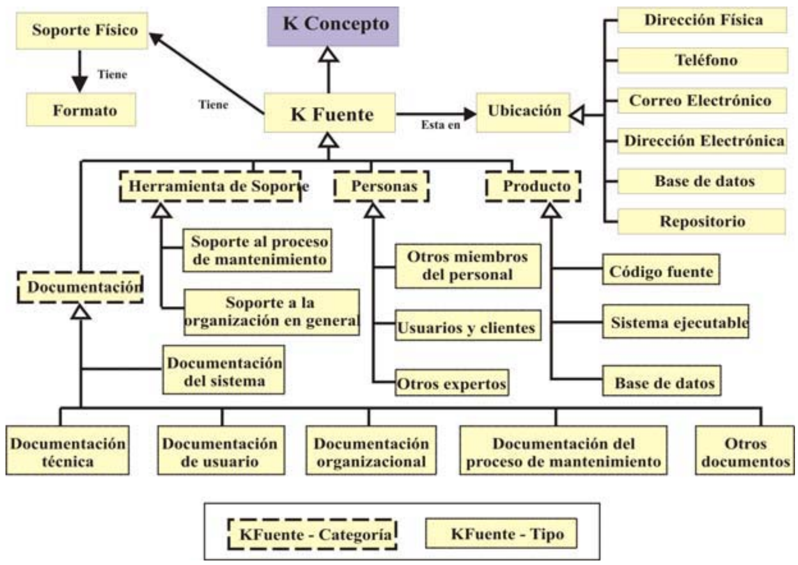


Fig. 1. Ontología de fuentes de conocimiento en el mantenimiento

Dentro de la categoría de **producto** podemos identificar tres clases principales: 1) *Sistema ejecutable*, 2) *Base de datos*, y 3) *Código fuente*.

Las **personas** constituyen una de las más importantes fuentes de conocimiento para los mantenedores, particularmente cuando el mantenedor no ha sido el desarrollador original de la aplicación [16]. Las personas a las que los ingenieros de mantenimiento suelen consultar han sigo agrupadas en las siguientes categorías:

1. *Usuarios y clientes*. Los usuarios son de gran ayuda a la hora de definir los requisitos que deben cubrir las modificaciones, así como para identificar las causas que originan los errores en el sistema al momento de corregirlos.
2. *Otros miembros del personal*. El apoyo de otros miembros del equipo resulta de gran ayuda, sobre todo cuando estos han sido los desarrolladores o han trabajado previamente con el sistema a mantener.
3. *Otros expertos*. En ocasiones los ingenieros del mantenimiento consultan a personas que no forman parte del personal, pero que son expertas de un dominio en específico, como por ejemplo en el manejo de cierto lenguaje o herramienta.

Las **herramientas** de soporte utilizadas por los mantenedores pueden ser muy variadas, por lo que resulta difícil establecer una categorización genérica. Entre este

tipo de herramientas se pueden encontrar las herramientas CASE tales como analizadores de código y de reingeniería inversa; sistemas de control de versiones y de control de configuración del software; sistemas de memorias organizacionales entre muchas otras mas.

Teniendo en cuenta el área en la que pueden dar soporte este tipo de herramientas, han sido clasificadas en las siguientes sub-categorías: 1) *Soporte al proceso de mantenimiento* y 2) S*oporte a la organización en general*.

Una vez identificadas las diferentes fuentes de conocimiento que los ingenieros del mantenimiento suelen consultar, se debe definir la forma en que el conocimiento será representado. Esta no es una tarea fácil, debido a que las fuentes de conocimiento son almacenadas en diferentes formatos y ubicadas en diferentes sitios.

Con el fin de resolver este problema se han añadido dos conceptos muy importantes para la ontología, los cuales son: *ubicación* y *soporte físico*. El concepto "ubicación" indica cómo puede ser localizada la fuente de conocimiento. Por ejemplo, si la fuente de conocimiento a consultar es una persona, es necesario conocer el correo electrónico, número de teléfono, dirección, etc., de la persona. El segundo concepto muestra el soporte físico y formato de la fuente de conocimiento. Por ejemplo, cada fuente puede tener uno o varios soportes físicos (libros, documentos electrónicos, videos, discos magnéticos, etc.), los cuales pueden estar en diferentes formatos (Word, pdf, Excel, dvd, VHS, etc). Además de esta, hemos creado otra ontología la cual define la relación entre los temas de conocimiento, sus fuentes, y actividades donde el conocimiento y fuentes son requeridos, generados o modificados. Esta ontología no ha podido ser explicada en este artículo por falta de espacio. Sin embargo, es importante aclarar que dicha ontología es necesaria para complementar la ontología de fuentes de conocimiento presentada y para ayudar en la recuperación de la información requerida para llevar a cabo cada actividad.

Ambas ontologías han sido completadas teniendo en cuenta la información obtenida en un caso de estudio donde un grupo de mantenedores de software fue estudiado [12]. Este trabajo nos permitió ver cómo cada tipo de ontología puede ayudar a reducir los problemas relacionados con la falta de conocimiento. Por ejemplo, ayudar a los ingenieros de mantenimiento a identificar las fuentes que tienen a mano, y qué conocimiento puede obtenerse de esta, con el fin de mejorar el flujo de conocimiento del grupo de mantenimiento. .

4 Conclusiones y trabajo futuro

La gestión del conocimiento es una técnica crucial para poder facilitar el trabajo de los mantenedores. Sin embargo, el primer paso para gestionar conocimiento es detectar las fuentes de conocimiento que existen en la organización y dónde pueden ser consultadas [13]. Con el fin de paliar este problema hemos presentado la ontología de fuentes de conocimiento que ayuda a identificar las fuentes de conocimiento usadas en una compañía de mantenimiento, así como la ubicación de las mismas.

Actualmente, estamos utilizando las ontologías propuestas como base para desarrollar un sistema de GC que recomiende la consulta de fuentes de conocimiento relacionadas con las tareas que lleven a cabo los ingenieros del mantenimiento. Para

implementar el sistema haremos uso de agentes inteligentes los cuales utilizarán dichas ontologías con el fin de detectar la información útil acorde a las necesidades de la tarea en la cual una persona trabaja. Esta herramienta será probada en varias compañías software con el fin de mejorar las ontologías y el propio sistema.

Referencias

1. Davenport, T.H., Long D. W. D., Beers, M.C.: Successful knowledge management projects. Sloan Management Review Winter (1998) 43-57
2. Dias, B., Anquetil, N., Oliveira, K..: Organizing the Knowledge Used in software Maintenance. Journal of Universal Computer Science. Vol. 9. (2003) 641-658
3. Gruber, T.: Towards Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human-Computer Studies. Vol. 43, (1995) 907-928
4. Kitchenham, B.A., Travassos, G.H., Mayrhauser, A., Niessink, F., Schneidewind, N.F., Singer, J.: Towards an Ontology of Software Maintenance. Journal of Software Maintenance: Research and Practice, Vol. 11, (1999) 365-389
5. Koskinen, J., Salminen, A., and Paakki, J.: Hypertext support for the information needs of software maintainers. Journal of Software Maintenance and Evolution: Research and Practice, Vol. 16, (2004) 187-215
6. Lethbridge, T.C., Singer, J., Forward, A.: How Software Engineerings Use Documentation: The State of the Practice. IEEE Software, Vol. 20. (2003) 35-39
7. Liebowitz, J.a.B., T.: Knowledge Organizations: What Every Manager Should Know. Washington: St. Lucie Press (1998).
8. Lindvall, M., and Rus, I.: Knowledge Management for Software Organizations. Managing Software Engineering Knowledge, A. Aurum, R. Jeffery, C. Wohlin, and M. Handzic (eds.). Springer, Berlin (2003) 73-94
9. IEEE Std 1219: Standard for Software Maintenance. USA, (1993)
10. Nebus, J.: Framing the Knowledge Search Problem: Whom Do We Contact, and Why Do We Contact Them? Academy of Management Best Papers Proceedings (2001) 1-7
11. Pfleeger, S.: Software Engineering: Theory and Practice (2001)
12. Rodriguez, O.M., Martínez, A.I., Favela, J., Vizcaíno, A., Piattini, M.: Undestanding and Supporting Knowledge Flows in a Community of Software Developers. In Groupware: Design, implementation, and Use, Proceedings of the X International Workshop on Groupware (CRIWG 2004). Springer, San Carlos, Costa Rica (2004) 52-66
13. Rodríguez, O.M., Martínez, A.I., Vizcaíno, A., Favela, J., Piattini, M.: Identifying Knowledge Flows in Communities of Practice. Encyclopedia of Communities of Practice in Information and Knowledge Management, E. Coakes and S.A Clarke (2005)
14. Ruiz, F., Vizcaíno, A., Piattini, M., García, F.: An Ontology for the Management of Software Maintenance Projects. International Journal on Software Engineering and Knowledge Engineering. Vol. 14, (2004) 323-346
15. Rus, I., Lindvall, M.: Knowledge Management in Software Engineering. IEEE Software. Vol. 19. (2002) 26-38
16. Seaman, C.: The Information Gathering Strategies of Software Maintainers. Proceedings of the International Conference on software Maintenance (2002) 141-149
17. Szulanski, G.: Intra-Firm Transfer of Best Practice Project: Executive Summary of the Findings. APQC (1994)

Representación del conocimiento basado en reglas para un diagnóstico enfermero

M. Lourdes Jiménez[1], José M. Santamaria[2], León A. González[1], Ángel L. Asenjo[3] y Luis M. Laita de la Rica[4]

[1] Departamento Ciencias de la Computación, ETSI Informática.
Universidad de Alcalá. 28871 Alcalá de Henares (Madrid).
{lou.jimenez, leon.gonzalez}@uah.es
http://www.cc.uah.es
[2] Gerencia Atención Primaria Área 11, Madrid.
jsantamaria.gapm11@salud.madrid.org
[3] Departamento de Enfermería, Escuela de Enfermería y Fisioterapia.
Universidad de Alcalá. 28871 Alcalá de Henares (Madrid).
angel.asenjo@uah.es
[4] Departamento de Inteligencia Artificial, Facultad Informática.
Universidad Politécnica de Madrid. 28660 Boadilla del Monte (Madrid).
laita@dia.fi.upm.es

Resumen. En este artículo se propone un Sistema Basado en Conocimiento para el diagnóstico del problema de salud del cansancio del rol del cuidador. Para construir la base de conocimiento de este Sistema ha sido necesario diseñar un Modelo del problema, dicha base de conocimiento está compuesta por un conjunto de reglas construidas en lógica bivaluada. El motor de inferencia de este Sistema utiliza las Bases de Gröbner y Formas Normales para obtener el diagnóstico desde la información contenida en la base de conocimiento. Para facilitar el acceso al Sistema se ha implementado un interfaz.

Palabras clave: Sistema Experto, Diagnóstico Enfermero, CoCoA, Representación del Conocimiento

1 Introducción

En este artículo se propone un Sistema Basado en Conocimiento para el diagnóstico enfermero del cansancio en el desempeño del rol del cuidador; diagnóstico estudiado por su importancia evidenciada por los cambios que la sociedad española viene sufriendo en los últimos años. El Sistema puede ser usado en consulta dado que muestra rápidamente el diagnóstico y además ofrece explicaciones de los resultados obtenidos, siendo de gran ayuda a los profesionales de la salud.

El Sistema implementado está formado por tres componentes básicos [5]:

- Una base de conocimiento que contiene el conocimiento de este problema.
- Un motor de inferencia que verifica la consistencia de la base de conocimiento y obtiene resultados automáticamente.
- Un interfaz de acceso e interacción para consultar dicho Sistema.

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 124-129, 2006.

Este es un Sistema Basado en Conocimiento que ha aplicado como técnica de representación del conocimiento reglas proposicionales construidas en lógica bivaluada, a partir de ahora se denotará como SBCBR. Así la base de conocimiento está formada por fórmulas lógicas proposicionales de la forma $\Phi \rightarrow \Psi$, donde Φ es una conjunción y/o disyunción de literales[1] y Ψ es un literal.

El motor de inferencia procesa la información expresada como fórmulas lógicas de la siguiente forma: primero, las fórmulas lógicas son traducidas a polinomios, y después, a dichos polinomios se les aplican las Bases de Gröbner [4] y las Formas Normales, usando el lenguaje de Álgebra Computacional CoCoA[2] (Computations in Commutative Algebra) [1], para comprobar la consistencia de la base de conocimiento y obtener consecuencias, en este caso, un diagnóstico.

El interfaz está implementado con el lenguaje de programación Java y se ha construido de forma que cualquier persona que desconozca temas como Álgebra Computacional, Lógica, Informática, etc. pueda usarlo, simplemente seleccionando algunos ítems para obtener un diagnóstico final sin necesidad de entender los procesos lógicos y algebraicos que son la base del SBCBR.

2 Base del conocimiento para el cansancio en el desempeño del rol del cuidador

La base de conocimiento que se propone está basada en un Modelo de conocimiento y se ha construido atendiendo al modelo disciplinar de D. Orem [2] y a la bibliografía diagnóstica. Tras una fase de adquisición de conocimiento llevada a cabo mediante toda la revisión bibliográfica, entrevistas con el experto, reuniones con profesionales de la salud, etc. se conceptualizaron y describieron dos "clases generales": la clase Núcleo de Análisis Inicial y la clase Diagnóstico. Dichas clases incluyen las siguientes "clases particulares" presentes en el diagnóstico del problema: Factores Condicionantes y Factores de Riesgo de la clase Núcleo de Análisis Inicial; Factores Etiológicos y Signos y Síntomas de la clase Diagnóstico. El siguiente paso permitió la definición de "subclases" de las clases Factores de Riesgo y Signos y Síntomas; y por último, la estructuración de sus "elementos concretos". El SBCBR se basa en el Modelo presentado en la Figura 1 y da respuesta a cinco diagnósticos distintos dentro del mismo problema.

2.1 Factores condicionantes básicos

Estos Factores son los que afectan a la persona, el cuidador, independientemente de si el problema está o no presente. Son sus características esenciales. Con las reglas construidas con estos factores se obtiene la Vulnerabilidad de la persona a padecer el problema. Estos Factores están relacionados con la edad, la posible limitación por parte del cuidador a realizar el cuidado adecuadamente, la responsabilidad que se

[1] Son variables proposicionales, que pueden estar precedidas por el símbolo $\neg$ (negación).

[2] Es un sistema para realizar cálculos en Álgebra Conmutativa.

tiene ante el cuidado, el nivel de instrucción y si realiza un trabajo fuera de casa.

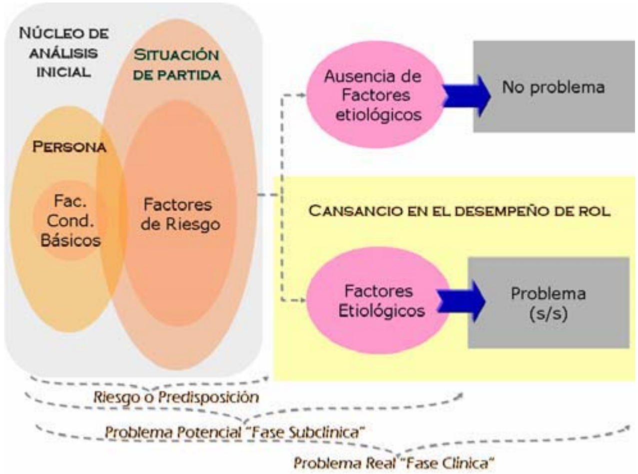


Fig. 1. Modelo del Diagnóstico.

A cada Factor Condicionante se le ha asignado una variable proposicional $x[i]$ ($i = 1,...,4$), precedida o no por el símbolo "$\neg$". Los Factores Condicionantes están unidos entre sí por la conjunción "y" ($\wedge$). Se asigna una "Intensidad" (denotada por "c") a cada conjunción de Factores Condicionantes. A mayor intensidad de los Factores Condicionantes mayor será la influencia en las intensidades de los Factores de riesgo y por lo tanto en el diagnóstico final. A los distintos grados de intensidades se les ha asignado los literales $c[i]$ con $i = 0,...,3$. (Véase Tabla 1). Este procedimiento se ha seguido con el resto de los factores y síntomas, obteniendo de este modo todas las reglas de producción del Sistema Experto.

¬x[1] (Común a todos)				
TC11	**x[4]∧x[5]**	**¬x[4]∧x[5]**	**x[4]∧¬x[5]**	**¬x[4]∧¬x[5]**
x[2]∧x[3]	3	3	3	2
x[2]∧¬x[3]	2	2	2	1
¬x[2]∧x[3]	2	1	1	1
¬x[2]∧¬x[3]	1	1	0	0

Tabla 1. Factores Condicionantes Básicos.

2.2 Factores de riesgo

Tienen sentido en la situación de partida, que es la situación en la que se encuentra una persona que está cuidando de otra, propiciando la presencia de Factores Etiológicos, la potenciación de los Factores Condicionantes y/o el agravamiento de la sintomatología del problema. Están divididos en las siguientes subclases: "Tipología del cuidado dependiente", "Cuidadora y Entorno" y "Relación Cuidadora y Cuidado".

2.3 Factores etiológicos

Los Factores Etiológicos hacen que el problema se active, siendo su presencia necesaria para considerar el problema potencial (fase preclínica).

2.4 Signos y síntomas

Son los datos que corroboran la presencia real del problema y su intensidad. Están divididos en las siguientes subclases: “Problema cuidador/a”, “Problemas de Relación con el Entorno” y “Problemas cuidado dependiente”.

3 Motor de inferencia

Una fórmula lógica A_0 es una consecuencia de las fórmulas $A_1,...,A_n$, que representan las reglas y hechos que componen la base de conocimiento del SBCBR, si siempre que $A_1,...,A_n$ son proposiciones verdaderas se tiene que A_0 también lo es.

Un paso muy importante en la teoría matemática que se utiliza en la construcción del SBCBR es el hecho de que en lugar de trabajar directamente con fórmulas lógicas, reglas y hechos se realiza una transformación de todas las reglas en polinomios; por lo que a partir de ese momento se trabajará con polinomios pasando así de operar con lógica bivaluada a operar con álgebra. Este proceso de traducir las fórmulas lógicas a polinomios es un proceso complejo y prolijo.

Una fórmula lógica A_0 es una consecuencia de las fórmulas $A_1,...,A_n$, que representan las reglas y hechos de un SBCBR, si y sólo si la traducción polinómica de la negación de A_0 pertenece al ideal generado por la traducción polinómica de las negaciones de $A_1,...,A_n$, junto con los polinomios que representan la lógica bivaluada $x_1^2 - x_1,...,x_n^2 - x_n$. Se denotará $NEG(A_0) \in J + I$, donde J es el ideal generado por las traducciones polinómicas de $\neg A_1,...,\neg A_n$ e I es el ideal generado por $x_1^2 - x_1,...,x_n^2 - x_n$.

3.1 Estudio de la consistencia

Un SBCBR es inconsistente si una contradicción es consecuencia de la información contenida en el SBCBR. Se puede demostrar que la inconsistencia se expresa con el hecho algebraico de que el elemento 1 del anillo de polinomios pertenece al ideal generado por los polinomios que traducen las negación de las reglas y hechos del SBCBR pues en este caso el ideal es el anillo entero. Si el ideal es todo el anillo, el teorema implicaría que todas las fórmulas son consecuencia del SBCBR, y en particular, las contradicciones también serían consecuencia del SBCBR. Así pues, el SBCBR es inconsistente si $1 \in J + I$.

4 Implementación del Sistema Experto

Usando el comando "NF" (Forma Normal) del sistema CoCoA se puede comprobar si un polinomio pertenece a un ideal y con el comando "ReducedGBasis" (Reduce la Base de Gröbner) se puede comprobar si el elemento unidad pertenece a un ideal. El proceso de implementación consta de las siguientes etapas:

- Traducción de las reglas y hechos a polinomios.
- Se obtiene el valor de la Base de Gröbner del ideal generado por esos polinomios.
- Traducir una pregunta A en una fórmula lógica y escribir la forma normal del polinomio que corresponde a la negación de dicha fórmula.

4.1 Implementación con el CoCoA

La herramienta CoCoA requiere que las fórmulas lógicas estén escritas en Forma Prefija. Los comandos negación, o lógico, y lógico e implicación se denotan a través de los símbolos $\neg, \vee, \wedge, \rightarrow$ respectivamente [4]. Una vez introducidas todas las reglas así como los hechos potenciales, es necesario comprobar la consistencia. El SBCBR es consistente si cada unión del conjunto de todas las reglas con un conjunto de hechos consistente no conduce a inconsistencia.

5 Interfaz

Se ha implementado un interfaz de usuario con la ayuda del lenguaje de programación Java. El SBCBR ejecuta CoCoA, pero el usuario no tiene por qué tener conocimientos de lógica y algebra para poder obtener un diagnóstico final. Gracias al interfaz el acceso al sistema es más intuitivo y agradable, haciendo totalmente transparente las operaciones matemáticas.

El usuario sólo tiene que contestar adecuadamente cada una de los formularios que la interfaz presenta. Al usuario se le presentaran 5 pantallas/ formularios distintos, y cada una de ellas informará del diagnóstico correspondiente: vulnerabilidad, riesgo, predisposición, potencialidad y severidad. (Ver Figura 2.)

6 Conclusiones

Se ha presentado un Sistema Basado en Conocimiento, construido mediante Reglas, de ayuda al diagnóstico del cansancio del rol del cuidador, además se ha representado el conocimiento de este diagnóstico a través de la definición de un Modelo de diagnóstico (Fig.1) que puede permitir, como novedad, la extrapolación a cualquier otra etiqueta diagnóstica, así como a otras áreas de conocimiento.

El Sistema SEDIEN será evaluado y valorado por un grupo de investigación for-

mado por profesionales del Centro de Salud Delicias del Área 11 de la Comunidad de Madrid, que estudian casos reales de pacientes que presentan este diagnóstico, el cual tiene una alta dificultad diagnóstica, por lo que SEDIEN será de gran ayuda una vez validado.

Como trabajo futuro, se está estudiando por una parte aplicar otro tipo de lógicas (polivalente, fuzzy, descriptiva, etc.) que se acerquen más al pensamiento humano abordando de este modo áreas en las que actualmente es difícil trabajar; y por otra la construcción de un segundo interfaz vía Web de modo que se pueda tener acceso a él vía Internet desde cualquier centro de salud.

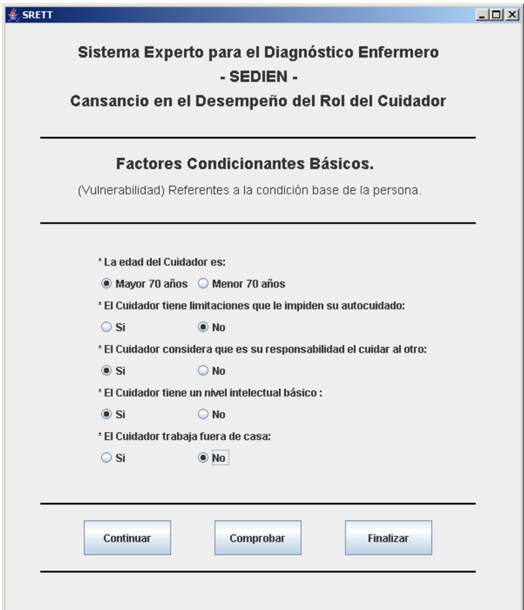


Fig. 2. Formulario de entrada de datos del SBCBR.

Referencias

[1] CoCoA, 2004 http://cocoa.dima.unige.it

[2] Orem D.E., 1993. Modelo de Orem: conceptos de enfermería en la práctica. Masson Salvat Enfermería.

[3] L. Jiménez, J.M. Santamaria et al., 2005. Ontology of the “fatigue in the performance of the caretaker roll”: a necessary step in the construction of a diagnosis expert system. IADIS Virtual Multi Conference on Computer Science and Information Systems

[4] T. Becker y V. Weispfenning, 1993. Gröbner bases. A computational approach to commutative algebra. Springer-Verlag.

[5] J. Giarratano y G. Riley, 2001. Sistemas expertos. Principios y programación. Internacional Thomson Editores, México.

Propuesta de un modelo de adquisición de habilidades y conocimiento complejo

Raquel Gilar Corbi y Juan Luis Castejón Costa

Universidad de Alicante
raquel.gilar@ua.es, jl.castejon@ua.es

Resumen. El objetivo principal de este trabajo es el de proponer y contrastar un modelo explicativo sobre la adquisición del conocimiento y las habilidades en el que se integren las principales teorías y modelos formulados hasta ahora sobre la adquisición de los conocimientos y las habilidades comprometidas en el desarrollo inicial de la competencia experta. A partir de aquí se plantean dos trabajos empíricos con el objetivo principal de delimitar los factores responsables de la adquisición del conocimiento y las habilidades en una situación de aprendizaje complejo como es la enseñanza superior universitaria. Los resultados obtenidos ponen de manifiesto que en la adquisición del conocimiento y las habilidades que conforman la expertez intervienen una serie de variables que contribuyen de forma conjunta a la explicación de los conocimientos adquiridos, como son la organización cualitativa del conocimiento, la habilidad intelectual, la motivación, el uso deliberado de estrategias y un ambiente rico de aprendizaje.

Palabras Clave: Modelos; Adquisición del Conocimiento; Habilidad Intelectual; Organización del Conocimiento

1 Introducción

El principal objetivo de este trabajo es proponer un modelo explicativo de la adquisición del conocimiento complejo y contrastar este modelo en varias situaciones instruccionales. La base teórica en la que se fundamenta este trabajo es el proceso de adquisición del conocimiento complejo, la adquisición de habilidades cognitivas, la inteligencia y el desarrollo de la competencia experta.

La investigación sobre la adquisición de habilidades cognitivas comenzó con el estudio de tareas simples como a resolución de puzzles, para posteriormente estudiar la resolución de problemas que pueden resolverse con la aplicación de un principio simple derivado de un conocimiento rico en un dominio determinado, y más recientemente, se ha centrado en la adquisición de grandes piezas de información (Beier & Ackerman, 2005; Kester, Lehnen, Van Gerven & Kirschner, 2006; Veenman, Elhout, & Meijer, 1997; Veenman, Wilhelm, & Beishuizen, 2004). Muchos de estos trabajos han examinado únicamente la adquisición de aspectos procedimentales de las habilidades. El siguiente paso sería examinar cómo los conceptos, los modelos mentales, y el conocimiento en un dominio específico se relacionan con los aspectos procedimen-

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 130-139, 2006.

tales del conocimiento, así como examinar sus relaciones con otras variables como la inteligencia, la motivación y las estrategias de adquisición del conocimiento. Además, debemos analizar cómo aprenden los estudiantes mediante formas más interactivas de instrucción, mediante tareas más reales en ambientes ecológicos de aprendizaje.

Basándonos en estas consideraciones proponemos un modelo que incluye los factores más importantes que forman parte de las teorías explicativas y modelos sobre la adquisición del conocimiento y habilidades implicadas en el desarrollo inicial de la competencia experta (Ericsson, 2005; Ericsson, Krampe, & Tesch-Römer, 1993; Sternberg, 1994, 1998a, 1999a). La característica fundamental del modelo es la interacción entre las variable, como se puede ver en la figura 1.

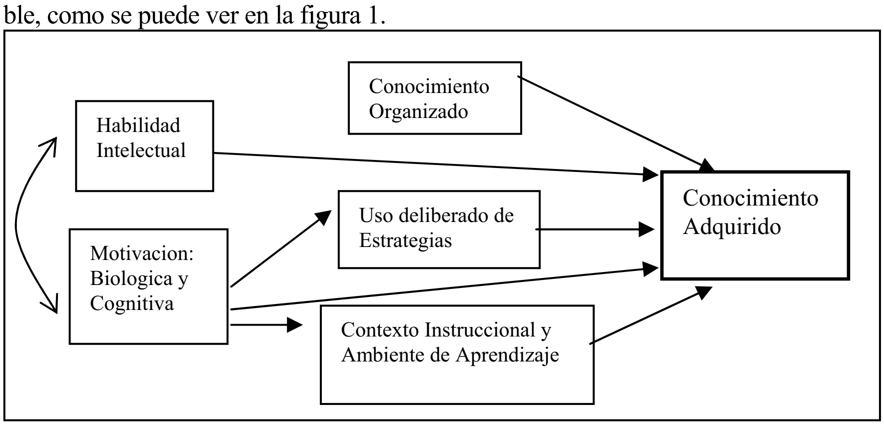


Fig. 1: Modelo propuesto de adquisición del Conocimiento y Habilidades

El término *habilidades* asume la existencia de varios aspectos de la capacidad intelectual que tienen sólo una moderada relación entre sí, a diferencia de la existencia de una única habilidad general representada por el factor g (Sternberg, 2003; Sternberg, Castejón, Prieto, Hautamäki, & Grigorenko, 2001).

Se reconoce un papel fundamental a la *Organización del Conocimiento* en el desarrollo de la competencia experta (Ericsson & Lehmann, 1996; Glaser, 1984, 1996; Patel, Kaufman, & Arocha, 2000), además de la habilidad intelectual. Esta habilidad para organizar el conocimiento es similar a las habilidades metacognitivas definidas por Veenman y colaboradores (Veenman & Elshout, 1999; Veenman & Spaans, 2005).

La *Motivacion* es, para muchos investigadores, el primer elemento necesario para la asdquisición de la competencia (Sternberg, 1999b). En el modelo de Sternberg (1999a) de desarrollo de la expertez, se entiende la motivación como auto-competencia.

El *uso deliberado de estrategias* durante el estudio y la práctica es un elemento clave de la teoría de la adquisición del conocimiento y habilidades implicadas en la competencia experta (Ericsson, Krampe & Tesch-Römer, 1993). (Veenman & Elshout, 1999; Veenman, Kok, & Blote, 2005).

El *contexto instruccional*. Determinados procedimientos instruccionales parecen estar relacionados más directamente con la adquisición de la competencia experta (Ericsson, et al., 1993). Goldman, Petrosino, y CTGV (1999) señalan la importancia de un ambiente rico de aprendizaje para la adquisición de la competencia experta.

Presentamos, a continuación, dos estudios en los que se pone a prueba el modelo propuesto de adquisición del conocimiento y habilidades.

2 Estudio 1

2.1 Método

Participantes

Los participantes en este trabajo son 110 estudiantes de primer curso de los estudios de Psicopedagogía de la Universidad de Alicante. Se trata de un grupo de estudiantes que posee conocimientos previos de carácter general de tipo psicológico y /o educativo.

Instrumentos y variables

a) Material didáctico. Consiste en una serie de temas que conforman el núcleo fundamental de la psicología de la instrucción, que forma parte del manual *Psicología de la instrucción* (Castejón, 2001).

b) Pruebas de evaluación de las características psicológicas

La prueba STAT (Sternberg Triarchic Abilities Test), nivel H, es un instrumento diseñado para evaluar los tres aspectos de la inteligencia triárquica, (Sternberg, 2000).

El instrumento utilizado para la evaluación de la motivación, mediante inventario, es el MAE (Motivación y Ansiedad de Ejecución) de Pelechano (1973b).

La evaluación de las estrategias de aprendizaje se realiza mediante el Cuestionario de Procesos de Estudio (CPE) –Study Process Questionnaire-, elaborado originalmente por Biggs (1987) con muestras de estudiantes universitarios.

La evaluación de los estilos de enseñanza y aprendizaje se realiza mediante el cuestionario de Estilos de Enseñanza- Aprendizaje (ESTIEA) (Castejón y Gilar, 2006).

c) Instrumento de evaluación de las estructuras cognitivas

La evaluación de la estructura del conocimiento adquirido durante el proceso de aprendizaje se realiza mediante el *pathfinder,* un procedimiento establecido por Schvaneveldt (1990). El pathfinder calcula varios índices, de los cuales, los más empleados son la medida de coherencia (COH) y la de similitud (SIM) de las matrices de proximidad de cada individuo o grupo. En nuestro trabajo se presentó a los estudiantes una matriz de 20 conceptos relativos al tema bajo estudio, en la que tenían que determinar la relación existente entre esos conceptos. El índice de similitud requirió una estructura referencial con la que comparar la de los estudiantes, que fue provista por dos miembros del equipo de investigación.

d) Evaluación del rendimiento final

La evaluación de los aprendizajes de los participantes en el trabajo se realizó mediante una prueba objetiva de rendimiento, que consistió en 20 enunciados con cuatro alternativas de respuesta, a los que los participantes tenían que responder con la alternativa correcta. La fiabilidad de consistencia interna de la prueba total fue de 0.70.

Procedimiento

En una primera fase, se procede a la aplicación de la prueba de evaluación de las habilidades intelectuales, el STAT, y la prueba de motivación general.

En la segunda fase, en los meses de noviembre a enero, se desarrolla el programa instruccional, se aplica la tarea de evaluación de conceptos, se aplica el Cuestionario de Procesos de Estudio, CPE, y el cuestionario ESTIEA.Al finalizar la fase instruccional, los participantes tuvieron que volver a cumplimentar de nuevo la tarea de evaluación de conceptos, en fase postest, en la misma sesión en la que se realizó la prueba de evaluación de conocimientos, y una vez finalizada la misma.

2.2 Resultados

Análisis correlacional

En este apartado se analizan los resultados de las correlaciones entre las 22 variables utilizadas. Los resultados de los coeficientes de correlación lineal de Pearson entre las variables se presentan en la tabla 1 junto con los estadísticos descriptivos, medias y desviaciones estándar, correspondientes a cada una de las variables.

Tabla 1.
Matriz de correlaciones entre las variables.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22
V1	1.00																					
V2	.47**	1.00																				
V3	.54**	.60**	1.00																			
V4	.05	.01	.02	1.00																		
V5	-.02	.15	.12	.04	1.00																	
V6	-.23	.05	-.15	.14	.28*	1.00																
V7	-.30*	-.01	-.02	.07	.31*	-.35*	1.00															
V8	.03	-.01	-.02	-.12	-.11	-.08	.04	1.00														
V9	.09	-.01	-.00	-.10	-.00	.05	-.11	-.18	1.00													
V10	.66	-.13	-.14	-.08	.01	.09	-.01	-.18	.60**	1.00												
V11	.04	-.02	-.15	-.19	-.22	-.03	-.13	.34**	.16	.17	1.00											
V12	-.06	.01	-.05	.00	-.00	.23	.18	.13	.52**	.44**	.19	1.00										
V13	.03	.03	.03	-.07	-.01	.11	.21	.42*	.10	.17	.31*	.35**	1.00									
V14	.04	.02	-.10	-.19	-.20	-.07	-.06	.80**	.00	.00	.83*	.18	.44**	1.00								
V15	.01	.00	-.03	-.05	-.00	.16	.04	-.02	.86**	.59**	.20	.89**	.26*	.12	1.00							
V16	.06	-.06	-.06	-.10	.00	.13	.14	.17	.44**	.74**	.32**	.51**	.79**	.30*	.55**	1.00						
V17	.10	.07	.01	.03	.07	.04	.03	.01	.24	.21	-.05	.16	.26*	-.03	.23	.31*	1.00					
V18	-.12	-.02	-.11	.05	.03	.13	.02	.29*	-.16	-.11	.20	-.08	.11	.30*	-.14	.00	-.01	1.00				
V19	.06	.08	.18	.06	.17	.10	.11	-.34**	.34**	.33**	-.05	.14	.02	-.23	.27*	.22	.29*	-.42**	1.00			
V20	.04	-.06	-.04	-.03	-.03	-.11	.05	.26*	-.06	-.04	-.12	.13	.46**	.23	.04	.29*	.40**	.15	.04	1.00		
V21	-.01	.35**	.21	.05	.16	.14	.15	-.07	-.05	-.10	-.14	-.01	.05	-.13	-.03	-.03	-.07	.04	-.03	-.11	1.00	
V22	.02	.31**	.26*	.08	.28*	.12	.35**	-.10	-.10	-.06	-.11	-.07	.06	-.13	-.09	.00	.08	-.13	.30*	-.03	.36**	1.00
Media	7.43	7.44	6.98	.49	.42	.24	.29	20.15	23.37	18.82	23.32	22.41	19.31	43.47	45.78	38.12	3.81	4.10	10.89	5.86	64.08	6.87
DS	2.11	2.41	2.79	.23	.20	.06	.09	3.49	4.02	4.74	3.74	4.33	4.96	5.99	7.18	7.41	2.58	1.77	2.32	2.04	15.40	1.46

*p= ó <.01; **p= ó <.001. V1= inteligencia analítica; V2= práctica; V3= creativa; V4= coherencia1; V5= coherencia2 V6= similitud1; V7= similitud2; V8= estrategia superficial; V9= estrategia profunda; V10= estrategia de logro; V11= motivo superficial; V12= motivo profundo; V13= motivo de logro; V14= acercamiento superficial; V15= acercamiento profundo; V16= acercamiento de logro; V17= sobrecarga de trabajo; V18= indiferencia laboral; V19= autoexigencia laboral; V20= Motivación positiva; V21= Preferencia por estilo de E/A;V22= rendimiento final.

Análisis de regresión paso a paso (stepwise)

En la tabla 2 se presentan los resultados del método paso a paso, utilizado para la predicción del rendimiento final que obtienen los participantes.

Tabla 2. Resultados del análisis de regresión realizado con el método paso a paso, tomando como criterio el conocimiento total adquirido.

R= .56; R^2 = .32; F= 10.67; Sign. F= .0000.

Variables en la ecuación

Variable	B	β	t	Sign. t
5	4.90	.28	3.28	.0014
2	.14	.21	2.28	.0244
17	.16	.25	2.90	.0046
19	.02	.22	2.39	.0186

Variables: 5= Similitud conceptual2; 2= Inteligencia práctica; 17= Autoexigencia en el trabajo/estudio; 19= Percepción ambiente de aprendizaje

Análisis causal mediante la técnica de ecuaciones estructurales

En la figura 2 se representa el modelo que mejor ajusta a los datos empíricos hallados en nuestro trabajo, con los valores de los parámetros estimados.

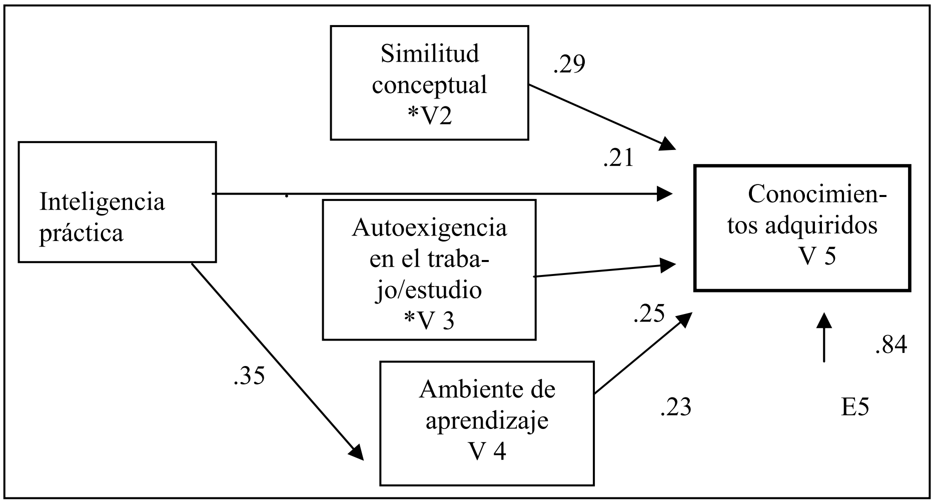


Fig. 2: Modelo de mejor ajuste a los datos acerca de los componentes de adquisición del conocimiento y las habilidades presentes en la competencia experta.

3 Estudio 2

3.1 Método

Participantes

Los participantes en este trabajo fueron 70 estudiantes de primer curso de los estudios de Psicopedagogía de la Universidad de Alicante.

Instrumentos y variables

En este estudio se emplearon los mismos instrumentos detallados en el estudio 1,

más los que pasamos a exponer a continuación.

- Material didáctico. Consiste en una unidad didáctica cuyo contenido está referido a la psicología del aprendizaje. El contenido de la unidad forma parte del manual *Psicología de la instrucción* (Castejón, 2001).

- Instrumentos de evaluación de la práctica, las estrategias de estudio y el aprendizaje independiente. El instrumento para la evaluación de estos aspectos es un diario, construido a tal efecto, en el que los estudiantes registran el tipo de actividad que llevan a cabo en cada momento del día, tanto las actividades generales, como las relativas a la práctica de estudio y de aprendizaje, relacionadas con el material a aprender, incluyendo las estrategias específicas empleadas durante el estudio de este material.

La sistematización de los datos recogidos en el diario se lleva acabo mediante un proceso sucesivo de elaboración inductiva de categorías a partir de las actividades específicas reseñadas por los participantes en el estudio. Este procedimiento es común en este tipo de estudios cualitativos (Miles & Huberman, 1994). Se derivaron 4 medidas en relación con las actividades de estudio y aprendizaje: el *uso de estrategias* (UESTRATE), la *frecuencia de uso de estrategias* (FESTRATE), el *tiempo de utilización de estrategias* (ESTRAVA) y el *tiempo total de estudio* (TIEMPOES).

Procedimiento

Los instrumentos se aplicaron siguiendo la misma secuencia que en el estudio 1, pero a lo largo de un mes en lugar de a lo largo de todo el curso.

Tabla 3.
Matriz de correlaciones entre las variables.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24
V1	1.00																							
V2	.32*	1.00																						
V3	.29	.45*	1.00																					
V4	-.06	.32*	.08	1.00																				
V5	-.16	.26	.19	.53*	1.00																			
V6	-.07	.02	.03	-.05	.09	1.00																		
V7	.13	-.05	.04	-.01	-.13	-.17	1.00																	
V8	.20	-.06	-.07	.00	-.11	-.27	.58*	1.00																
V9	-.07	-.03	-.26	.08	-.01	.30	.11	.13	1.00															
V10	-.07	.06	.09	.01	-.12	.13	.53*	.28	.13	1.00														
V11	.04	.24	.12	.02	.14	.36*	.22	.20	.25	.30	1.00													
V12	-.09	-.01	-.15	.03	.04	.76*	-.03	-.06	.84*	.17	.37*	1.00												
V13	.03	.00	.07	.00	-.14	-.02	.87*	.49*	.14	.87*	.30	.08	1.00											
V14	.15	.13	.03	.02	.02	.08	.50*	.73*	.25	.38*	.81*	.22	.50*	1.00										
V15	.24	.16	.08	.18	.06	.02	.20	.17	-.04	.14	.44*	-.01	.19	.41*	1.00									
V16	-.19	-.13	-.11	.16	.14	.35*	-.17	-.09	.21	-.06	.09	.34	-.14	.01	.06	1.00								
V17	-.18	.07	.09	.27	.21	-.20	.20	.02	.04	.01	-.00	-.08	.12	.01	.27	-.19	1.00							
V18	.04	.08	-.11	-.07	.02	.19	.02	-.02	.14	.09	.42*	.20	.07	.27	.37*	.15	-.08	1.00						
V19	.29	.16	.22	.02	.15	.04	.23	.00	.11	.20	.18	.09	.25	.12	-.07	-.09	-.13	.20	1.00					
V20	.16	-.09	-.06	-.25	-.10	.07	-.08	.00	-.02	-.14	-.09	.02	-.12	-.05	.08	.02	.09	-.07	-.09	1.00				
V21	.14	.13	-.03	.11	.06	-.08	.10	.19	.14	.08	.04	.04	.11	.14	.12	-.29	.26	.05	.05	.24	1.00			
V22	.10	-.03	-.10	-.16	-.08	.00	-.09	.02	.05	.00	-.18	.03	-.05	-.11	-.09	-.16	-.06	-.23	-.25	.55*	.49*	1.00		
V23	.15	-.12	-.15	-.19	-.11	.07	-.11	.03	.02	-.10	-.20	.06	-.12	-.12	-.02	-.01	-.06	-.19	-.23	.84*	.29	.85*	1.00	
V24	.10	.40*	.22	.40*	.62*	-.08	-.09	-.17	.12	.01	.09	.03	-.04	-.03	.15	-.03	.34*	.13	.35*	.06	.31*	-.03	-.14	1.00
Media	7.39	7.48	6.66	.46	.33	20.71	22.60	18.48	23.42	22.90	20.32	44.13	45.50	38.81	3.68	4.37	10.62	6.15	64.68	13.04	.24	.59	3.33	7.25
DS	1.65	2.00	2.18	.18	.08	3.49	4.13	4.49	4.12	4.21	5.32	6.24	7.21	7.65	2.79	1.79	2.21	2.13	6.86	14.26	.09	.37	3.73	1.29

*p= ó <.01. V1= inteligencia analítica; V2= práctica; V3= creativa; V4= coherencia2; V5= similitud2; V6= estrategia superficial; V7= estrategia profunda; V8= estrategia de logro; V9= motivo superficial; V10= motivo profundo; V11= motivo de logro; V12= acercamiento superficial; V13= acercamiento profundo; V14= acercamiento de logro; V15= sobrecarga de trabajo; V16= indiferencia laboral; V17= autoexigencia laboral; V18= Motivación positiva; V19= Preferencia por estilo de E/A; V20= tiempo de estudio; V21= uso estrategias; V22= frecuencia uso estrategias; V23= tiempo uso estrategias; V24= calificación final

3.2 Resultados

Análisis correlacional

En este apartado se analizan los resultados de las correlaciones entre las 24 variables utilizadas. Los resultados de los coeficientes de correlación lineal de Pearson

entre las variables se presentan en la tabla 3.

Análisis de regresión múltiple con el método paso a paso

En la tabla 4 se presentan los resultados del método paso a paso, utilizado para la predicción del rendimiento final que obtienen los participantes.

En la figura 3 se representa el modelo que mejor ajusta a los datos empíricos hallados en nuestro trabajo, con los valores de los parámetros estimados.

Tabla 4. Resultados del análisis de regresión realizado con el método paso a paso, tomando como criterio el conocimiento total adquirido.

R= .77; R^2 = .60; F= 16.19; Sign. F= .0000.

Variables en la ecuación

Variable	B	β	T	Sign. T
5	7.46	.47	5.16	.0000
21	2.73	.18	2.05	.0451
19	.05	.27	3.01	.0040
17	.12	.21	2.26	.0279
2	.12	.19	2.11	.0390

N= 59. Variables: 5= Similitud conceptual2; 21= Uso deliberado de estrategias; 19= Percepción ambiente de aprendizaje; 17= Autoexigencia en el trabajo/estudio; 2= Inteligencia práctica.

Análisis causal mediante la técnica de ecuaciones estructurales

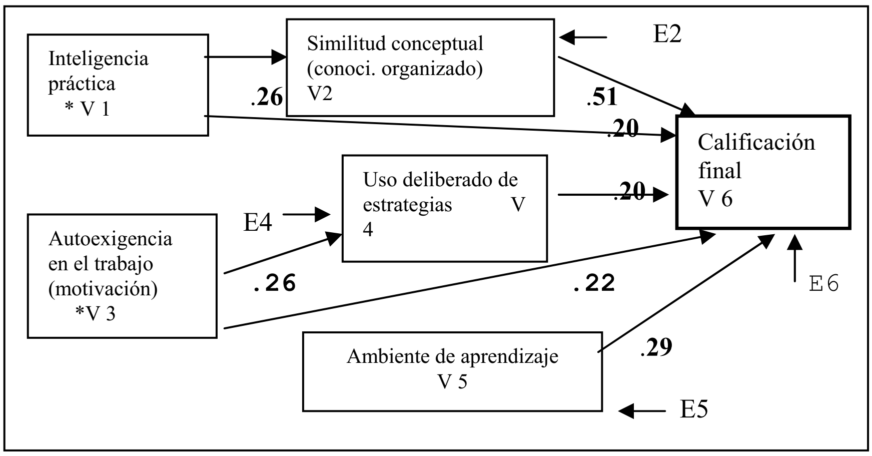


Fig. 3: Modelo estructural sobre las relaciones entre los componentes de adquisición del conocimiento, que mejor ajusta a los datos empíricos.

4 Discusión

Las variables que muestran una relación significativa con los conocimientos y habilidades adquiridas durante en proceso de enseñanza/aprendizaje de un material significativo, complejo, llevado a cabo en una situación educativa real, giran alrededor de los aspectos de la *habilidad intelectual*, la *organización del conocimiento*, la *motivación*, el *uso de estrategias de aprendizaje*, y la *percepción del contexto instruccional* en que se lleva a cabo este proceso. Estos aspectos son precisamente los que en mayor o menor medida se encuentran presentes en las teorías, modelos e hipótesis explicativas de la adquisición del conocimiento y las habilidades que forman parte del desarrollo de la competencia experta (Ericsson & Lehmann, 1996; Sternberg, 1994, 1998,a, 1999,a).

Los resultados relativos a la *habilidad intelectual* ponen de manifiesto que el tipo de inteligencia que aparece relacionado con la adquisición de conocimientos es el de la inteligencia práctica.

La *calidad de la organización del conocimiento*, definida de forma operativa por las variables de coherencia y similitud conceptual, es el elemento que mayor influencia tiene sobre la adquisición del conocimiento y las habilidades. De esta forma se reconoce el papel predominante que tiene el conocimiento en el desarrollo de la competencia experta (Beier y Ackerman, 2005; Patel, Kaufman, & Arocha, 2000; Veenman, Kok y Blote, 2005; Vincent, Decker & Munford, 2002), independientemente de las habilidades intelectuales. No obstante, el hecho de que las habilidades intelectuales también ejerzan un efecto directo sobre el conocimiento e indirecto a través de éste, sobre la adquisición de los conocimientos y habilidades que forman parte de la competencia experta, está también de acuerdo con la teoría sintética de la expertez (Sternberg, 1994, 1999,a) y la teoría de la complejidad cognitiva de Ceci (1996) sobre la adquisición de la competencia, que destacan el carácter interactivo de ambos elementos.

El *uso deliberado de estrategias* durante el estudio es otra de las variables que tiene una influencia directa sobre la adquisición del conocimiento. Ninguna de los factores del cuestionario de procesos de estudio (CPE), muestra una relación significativa de orden cero con los resultados de aprendizaje. Nuestros resultados muestran que el uso real y deliberado de estrategias es lo que influye la adquisición del conocimiento, más que el tiempo total dedicado al estudio. Estos resultados están de acuerdo con la teoría de la práctica deliberada, según la cual es el tipo de actividad realizada de forma deliberada, consciente y con esfuerzo, lo que tiene efectos positivos sobre el desarrollo de la competencia (Ericsson, et al., 1993; Ericsson & Lehmann, 1996).

La *motivación* es otro de los factores que incide tanto de forma directa como indirecta sobre los resultados de adquisición del conocimiento y las habilidades. La motivación es un mecanismo complejo en el que intervienen factores biológicos y cognitivos (Covington, 2000) que determinan el impulso general a la actividad, la motivación hacia el logro o el sentimiento de autoeficacia.

Un elemento destacado del modelo es *el contexto* en el que se desarrolla la competencia. La preferencia por un ambiente rico y variado de aprendizaje está relacionada positivamente con la adquisición del conocimiento y las habilidades. La implicación instruccional para el desarrollo de la competencia experta parece clara, se deben de

favorecer ambientes ricos de aprendizaje que estimulen la adquisición de dicha competencia (Goldman et al., 1999, Nokes y Ohlsson, 2005).

En suma, en nuestro trabajo se han identificado un conjunto de variables que están directamente implicadas en el desarrollo inicial de la competencia experta, y se ha establecido la forma precisa en la que estas variables se relacionan entre sí, dentro de un modelo que tiene en cuenta las principales hipótesis explicativas formuladas sobre la adquisición de la competencia.

Referencias

Beier, M.E., y Ackerman, P.L. (2005). Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge: Promising results in a real-world learning environment. Psychology and Aging, 20(2), 341-355.

Biggs, J. B. (1987). Study Process Questionnaire (SPQ). Hawthorn, Victoria: Australian Council for Educational Research.

Castejón, J. L. (2001). Introducción a la psicología de la instrucción . Alicante, Spain: Ediciones Club Universitario.

Castejón, J.L. y Gilar, R. (2006). Evaluación del estilo de enseñanza-aprendizaje en estudiantes universitarios. Revista de Psicología y Educación (en prensa).

Ceci, S. (1996). On intelligence. A bioecological treatadise on intellectual development. (Expanded edition). Cambridge, MA: Harvard University Press.

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. Annual Review of Psychology, 51, 171–200.

Ericsson, K.A. (2005). Superior decision making as an integral quality of expert performance: Insights into the mediating mechanisms and their acquisition through deliberate practice. In R. Lipshitz y H. Montgomery (Eds.), How professional make decisions (pp. 135-167). Mahwah: Laurence Erlbaum Associates.

Ericsson, K. A., Krampe, R. T., y Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. Psychological Review, 100, 363–406.

Ericsson, K. A., y Lehmann, A. C. (1996). Expert and exceptional performance: Evidence on maximal adaptations on task constraints. Annual Review of Psychology, 47, 273–305.

Glaser, R. (1984). Education and thinking. The role of knowledge. American Psychologist, 39(2), 93–104.

Glaser, R. (1996). Changing the agency for learning: Acquiring expert performance. In K. A. Ericsson (Ed.), The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games (pp. 303–311). Hillsdale, NJ: LEA.

Goldman, S.R., Petrosino, A.J., y Cognition and Technology Group at Vanderbilt (1999). Design principles for instruction in content domains: Lessons from research on expertise and learning. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay, & M. T. H. Chi (Eds.), Handbook of Applied Cognition (pp. 595–627). New York: John Wiley & Sons.

Kester, L., Lehnen, C., Van Gerven, P., y Kirschner, P. (2006). Just-in-time, schematic supportive information presentation during cognitive skill acquisition. Computers in Human Behavior, 22, 93-112.

Miles, M., y Huberman, M. (1994). Qualitative data analysis. 2nd edition. Thousand Oaks, CA: Sage Publications.

Nokes, T., y Ohlsson, S. (2005). Comparing multiple paths to mastery: What is learned?. Cognitive Science, 29(5), 769-796.

Patel, V. L, Kaufman, D. R., y Arocha, J. F. (2000). Conceptual change in the biomedical and health sciences domain. In R. Glaser (Ed.), Advances in instructional psychology: Educational design and cognitive science. Vol 5 (pp. 329–392). Mawhah, NJ: LEA.

Pelechano, V. (1973b). Manual del cuestionario MAE Madrid, Spain: Fraser.

Schvaneveldt, R. W. (1990). Pathfinder associative networks. Studies in knowledge organization. Norwood, NJ: Ablex Publishing Co.

Sternberg, R. J. (1994). Cognitive conceptions of expertise. International Journal of Expert Systems, 7(1), 1–12.

Sternberg, R. J. (1998a). Abilities are forms of developing expertise. Educational Researcher, 27(3), 11–20.

Sternberg, R. J. (1999a). Intelligence as developing expertise. Contemporary Educational Psychology, 24, 359–375.

Sternberg, R. J. (1999b). Ability and expertise. It's time to replace the current model of intelligence. American Educator, Spring, 10–13 and 50–51.

Sternberg, R. J. (2000). The concept of intelligence. In R. J. Sternberg (Ed.), Handbook of intelligence (pp. 3–15). New York: Cambridge University Press.

Sternberg, R. J. (2003). Construct validity of the theory of special intelligence. In R. J. Sternberg, J. Lautrey, & T. I. Lubart (Eds.), Models of intelligence: International perspectives (pp. 55–77). Washington: American Psychological Association.

Sternberg, R. J., Castejón, J. L., Prieto, M. D., Hautamäki, J., y Grigorenko, E. (2001). Confirmatory factor analysis of the Sternberg Triarchic Abilities Test (Multiple Choice Items) in three international samples: An empirical test of the triarchic theory. European Journal of Psychological Assessment, 17, 1–16.

Veenman, M., y Elshout, J. J. (1999). Changes in the relation between cognitive and metacognitive skills during the acquisition of expertise. European Journal of Psychology of Education, XIV, 4, 509–523.

Veenman, M., Elshout, J. J., y Meijer, J. (1997). The generality vs domain-specificity of metacognitive skills in novice learning across domains. Learning and Instruction, 7(2), 187–209.

Veenman, M., Kok, R., y Blote, A. (2005). The relation between intellectual and metacognitive skills in early adolescence. Instructional Science, 13, 193-211.

Veenman, M., y Spaans, M. (2005). Relation between intellectual and metacognitive skills: Age and task differences. Learning and Individual Differences, 15, 159-176.

Veenman, M., Wilhelm, P., y Beishuizen, J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. Learning and Instruction, 14(1) 89-109.

Vincent, A. S., Decker, B. P., y Munford, M. D. (2002). Divergent thinking, intelligence, and expertise: A test of alternative models. Creativity Research Journal, 14(2), 163–178.

Razonamiento temporal en una aplicación de gestión de enfermería

J. Salort, J. Palma y R. Marín

Artificial Intelligence and Knowledge Engineering Group
University of Murcia, Spain
salort@um.es

Resumen Presentamos una tarea de investigación en curso consistente en conectar una aplicación de gestión de hojas de enfermería hospitalarias con (i) una ontología temporal general, de modo que la aplicación de enfermería almacene y recupere la componente temporal de la información que maneja en la ontología en vez de en la base de datos, y (ii) con un razonador temporal general, de modo que un usuario pueda lanzar una pregunta compleja con fuerte componente temporal sobre la situación en enfermería o sobre algún paciente y reciba informes detallados que respondan a su pregunta. La aplicación de enfermería está pendiente de implantación en un entorno hospitalario real.

Palabras clave: Inteligencia artificial: Ingeniería del conocimiento, Interfaces inteligentes, Ontologías, Resolución de problemas, Sistemas expertos.

1. Introducción

La aplicación de gestión de hojas de enfermería en hospital es una aplicación de gestión normal, pero que destaca por el tratamiento que se le da a la información temporal que almacena. Su contribución científica es que va a almacenar algunos datos en una ontología en vez de en base de datos y que va a trabajar conectada con un razonador temporal, y parte de la información que muestre serán respuestas complejas a preguntas complejas de carácter temporal formuladas gráficamente o textualmente por el usuario. El razonador temporal es un trabajo del grupo de investigación ya terminado [5]. La ontología temporal va mejorando según la vamos perfeccionando [11]. Y la aplicación de gestión de enfermería está pendiente de las ampliaciones y cambios que sugieran los médicos que trabajan con ella de prueba.

La representación y el razonamiento sobre el tiempo juegan un papel importante en diversas áreas de la inteligencia artificial [15], tales como procesamiento del lenguaje natural, planificación, scheduling, diagnóstico, informática en medicina [2] o minería de datos temporales. En estas tareas, la información temporal que queremos representar puede ser cualitativa o

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 140-145, 2006.

cuantitativa. Es necesario decidir cuáles son las entidades temporales básicas, y hay dos posibilidades clásicas que a menudo han sido objeto de discusión [1]: puntos temporales (es decir, instantes) o intervalos temporales (es decir, eventos). Íntimamente ligado con esto está el problema de decidir el conjunto de relaciones primitivas entre los objetos temporales. También es necesario decidir la estructura del tiempo, que puede ser acotado o ilimitado, discreto o denso, y la clase de orden (total, parcial, ramificado, circular) [3]. Además podemos decidir representar un intervalo de tiempo como un conjunto borroso fuzzy y así generalizar las relaciones temporales cualitativas, considerando instantes e intervalos de tiempo borrosos [7].

2. Arquitectura

La arquitectura del sistema es semejante a la de cualquier sistema de aplicaciones heterogéneas conectadas entre sí mediante interfaces de servicios web. En la Figura 1 vemos su estructura. La parte central del sistema es la aplicación de enfermería manejada por un usuario, que es la que lleva el control.

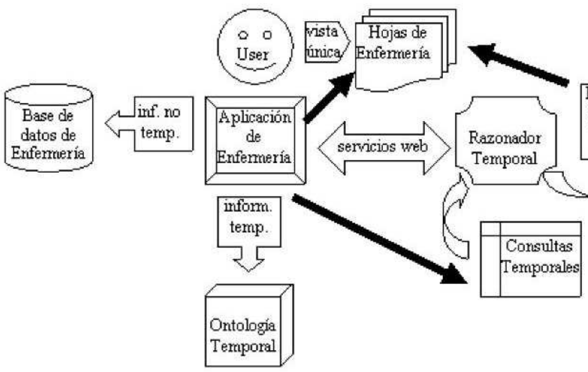


Figura 1. Arquitectura para la conexión de la aplicación de enfermería con la ontología temporal y con el razonador temporal. Al usuario se le oculta la existencia del razonador temporal, cuyas entradas son generadas automáticamente por la aplicación en respuesta a acciones del usuario, y cuyas salidas aparecen integradas visualmente en las hojas de enfermería.

La información no temporal se guarda en una base de datos relacional que tiene soporte para transacciones, siguiéndose en esto el esquema clásico tripartito de *presentación–aplicación–datos*, y donde la parte de presentación visual que maneja el usuario se construye según el esquema de tres capas *modelo–vista–control*.

La información temporal se guarda en un servidor de ontologías [8] que permite almacenar información temporal estructurada en forma de árbol y conectada con los conceptos del dominio médico de enfermería (los *eventos*

de enfermería serán *instantes temporales* y los *estados de enfermería* serán *intervalos temporales*).

La hoja de enfermería envía consultas al razonador temporal y recibe los informes de respuesta mediante servicios web, que aún están sin definir pero que se concretarán en un lenguaje de consultas y respuestas temporales construido sobre OWL [16]. En vez de diseñarlo nosotros directamente, estamos estudiando la posibilidad de extender subconjuntos de otros lenguajes ya definidos para la web semántica (GLIF3, Asbru) [4,12] que manejan conceptos temporales de forma rudimentaria y mejorable en nuestra opinión.

De este modo, externamente el usuario médico o enfermero puede manejar todo tipo de información (temporal y no temporal), pero observa que la aplicación es muy potente en lo relativo al tiempo (consultas e informes), pues es capaz de procesar cuestiones temporales en un nivel de abstracción llamativamente alto.

3. La ontología temporal

Ya hemos terminado la construcción de la primera versión de una ontología temporal general en OWL con soporte de eventos, estados, e históricos de ambos. Incluye soporte para los tres tipos de entidades y relaciones temporales (entre instantes, intervalos, y duraciones), y permite su especificación mediante números difusos (fuzzy), lo cual no se había hecho previamente en ninguna ontología temporal [9]. Trabaja con referencias deícticas (como *hoy, ahora*) y permite expresar conceptos temporales vagos (con *aproximadamente*).

En nuestra ontología existirán tres tipos de entidades temporales, interrelacionadas entre sí: *intervalos, instantes, duraciones*. Dos instantes, o un instante y una duración, implicarán un único intervalo, el cual a su vez implicará una única duración; esta relación de implicación aparecerá explícitamente representada en la ontología, de forma que podremos navegar desde unas entidades hasta otras (mediante enlaces representados como propiedades de las clases de la ontología). Un intervalo puede ser no acotado por la izquierda o derecha cuando no tiene principio o fin (ese instante estaría en el infinito). Nuestros tipos de datos primitivos son seis: *año, mes, día, hora, minuto, segundo*. Estos tipos primitivos pueden utilizarse de dos maneras: para especificar constantes temporales en el razonador temporal, y para especificar duraciones temporales en el razonador temporal.

4. La aplicación de enfermería

Nuestra aplicación de Gestión de Hojas de Enfermería opera con salidas y entradas fisiológicas del paciente, sus constantes vitales y sus gráficas asociadas, tratamientos administrados, así como parámetros respiratorios y analítica (datos de laboratorio). La aportación principal es la búsqueda avanzada por campos sensible al tipo de información desde PDA, Web, o aplicación. Además, los datos recogidos nos permitirán en un futuro realizar minería de datos y evaluación de

la calidad asistencial con la información introducida por los médicos en la base de datos. La aplicación está integrada con CH4 (el sistema de información de historias clínicas también desarrollado por el grupo) [10].

El historial del paciente en enfermería es visto como un conjunto de informes u hojas, en cuyo modelo se puede configurar hasta el nivel de tabla, fila, columna, y celda, con cadenas, enteros, y fechas, cada uno de los campos o tablas de la hoja, de modo que en cada hospital puedan personalizar las hojas según se decida en cada Comisión de Historias Clínicas (entidades que definen en cada hospital el formato de los documentos de historia clínica del paciente, incluyendo las hojas de enfermería).

5. Conexión con el razonador temporal

En el razonador temporal [5,15] es posible construir expresiones temporales con conjunciones, disyunciones, y negaciones. Las expresiones pueden hacer referencia a variables o a constantes temporales compuestas por agregación de varias unidades temporales (segundo, minuto, hora, día, mes, año). Una expresión consistirá entonces en un árbol de sub-expresiones unarias (negación) o binarias (conjunción y disyunción), cuyas hojas son siempre expresiones atómicas. Una expresión atómica consiste en la comparación de una variable con una constante utilizando algún operador válido (mayor, menor, igual, o dos de estos juntos). Una expresión atómica también puede consistir en el modificador *aproximadamente*, que se aplica sobre otra expresión atómica y que permite cierta tolerancia en las comparaciones. Sobre estas expresiones podemos construir preguntas de posibilidad (*may*) y necesidad (*must*). La conexión del razonador temporal con la ontología temporal parece ser por todo ello un trabajo abordable en un tiempo no excesivamente largo.

Las relaciones temporales establecerán precedencias entre las entidades. Estas entidades primitivas se relacionan mediante restricciones cualitativas o cuantitativas. Las relaciones cualitativas pueden ser entre punto y punto (tres relaciones), entre punto e intervalo (cinco relaciones), entre intervalo y punto (cinco relaciones), y entre intervalo e intervalo (las trece relaciones de Allen [1]). Las relaciones cuantitativas relacionan punto y punto, o bien duración y duración, y vienen expresadas con un número fuzzy. Ambos tipos de relaciones serán binarias. La literatura sobre razonamiento temporal es extensa en estos aspectos, y no es necesario que nos detengamos en ella por ser bien conocida [15,1,3,7].

Para la conexión de la aplicación de enfermería con el razonador temporal mediantes servicios web, usaremos el lenguaje OWL, que hoy en día se reconoce como uno de los mejor adaptados a los requerimientos de la web semántica. Los detalles de esta conexión y la forma en que se hará en particular, están aún sin concretar y pendientes de que se tome una decisión al respecto.

En la Figura 2 vemos dos ejemplos de consultas temporales que pueden hacerse, traducidas a un lenguaje previo al lenguaje que entiende el razonador temporal, más apto para la inspección humana.

```
may
    [(constantes : F.R. <= 115)   or   (constantes : F.C. < 180)]
before
    [not   [(constantes : T.a. <= 37,5)   and   (swan ganz : P.S.   in   {"low", "med"})]]
approx   15   min

  must
      [(constantes : T.A.max. >= 25)   and   (swan ganz : P.D. = "100 %")]
  started-by
      [(swan ganz : P.M. ! = "yes")   or   [not   (swan ganz : C.P.V. >= 1,02)]]
  between   5   and   15   sec
```

Figura 2. Dos ejemplos de consultas temporales. La primera es una consulta de posibilidad, donde se pregunta si un suceso puede ocurrir aproximadamente quince minutos antes que otro suceso. Los sucesos son conjunciones y disyunciones y negaciones de átomos consistentes en comparaciones de un valor de la hoja de enfermería con una constante. Los sucesos van entre corchetes, los átomos van entre paréntesis, y los comparadores son los signos matemáticos que ya conocemos (<, >, =, !) junto con (*in*). Las constantes pueden ser numéricas o cadenas.

6. Conclusiones y vías futuras

La conclusión principal es que nuestra propuesta combina en un modelo general las tres aportaciones (aplicación de gestión de enfermería, ontología temporal, razonador temporal) que aparecen por separado en otras propuestas previas de otros autores, y añade algunas innovaciones, con el objetivo de conseguir un modelo reusable en distintos dominios, que sea lo más general posible. Todas las aportaciones y ventajas expuestas a lo largo de este artículo justifican el valor añadido de la integración de un razonador temporal y de un servidor ontologías con una aplicación de gestión de hojas de enfermería hospitalarias en entornos de UCI y planta.

Como vías futuras, se están realizando numerosas visitas al Hospital General de Elche porque allí tenemos investigadores médicos, con ayuda de los cuales hemos conseguido construir el sistema de gestión de hojas de enfermería para UCI y planta. En la actualidad estas visitas están siendo extendidas al Hospital Universitario de Getafe, donde es de esperar que podamos firmar un contrato para implantar el sistema, y con la clínica Virgen de la Vega de Murcia (ASISA). Podemos apuntar ya, como trabajo futuro, a la explotación que se puede hacer de estos desarrollos en el marco de herramientas para el control de la calidad asistencial [6,13,14].

Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Educación y Ciencia bajo el proyecto MEDICI (TIC2003-09400-C04-01).

Referencias

1. J. Allen (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM,* 26(11):832–843.
2. E. Antman, D. Anbe, P. Armstrong, E. Bates, L. Green, M. Hand, J. Hochman, H. Krumholz, F. Kushner, G. Lamas, C. Mullany, J. Ornato, D. Pearle, M. Sloan, S. Smith (2004). *ACC/AHA guidelines for the management of patients with ST–elevation myocardial infarction.* American College of Cardiology/American Heart Association Task Force on Practice Guidelines.
3. P. van Beek (1991). Temporal query processing with indefinite information. *Artificial Intelligence in Medicine,* 3:325–339.
4. A. Boxwala, M. Peleg, S. Tu, O. Ogunyemi, Q. Zeng, D. Wang, V. Patel, R. Greenes, E. Shortliffe (2004). GLIF3: A Representation Format for Sharable Computer–Interpretable Clinical Practice Guidelines. *Journal of Biomedical Informatics,* 37:147–161.
5. M. Campos, A. Cárceles, J. Palma, R. Marín (2002). A General Purpose Fuzzy Temporal Information Management Engine. Workshop on Formal Modeling of Intelligent Peripheral Systems, *1st Eurasian Conference on Advances in Information and Communication Technology* (EURASIA–ICT 2002).
6. P. de Clercq, J. Blomb, H. Korsten, A. Hasman (2004). Approaches for creating computer–interpretable guidelines that facilitate decision support. *Artificial Intelligence in Medicine,* 31:1–27.
7. D. Dubois, H. Prade (1989). Processing fuzzy temporal knowledge, *IEEE Transactions on Systems, Man and Cybernetics,* 19(4):729–744.
8. A. Gómez–Pérez, M. Fernández–López, O. Corcho (2004). *Ontological Engineering.* Springer.
9. J. R. Hobbs, F. Pan (2004). An Ontology of Time for the Semantic Web. *ACM Transactions on Asian Language Information Processing,* 3:66–85.
10. J. M. Juárez, J. T. Palma, M. Campos, J. Salort, A. Morales, R. Marín (2005). A model–based architecture for fuzzy temporal diagnosis. *Tenth International Conference on Computer Aided Systems Theory* (EUROCAST 2005).
11. J. Salort, J. Palma, R. Marín (2005). Una ontología temporal general para la Web semántica. IV Workshop on Planning, Scheduling and Temporal Reasoning, *XI Conference of the Spanish Association for Artificial Intelligence* (CAEPIA 2005).
12. Y. Shahar, O. Young, E. Shalom, M. Galperin, A. Mayaffit, R. Moskovitch, A. Hessing (2004). A framework for a distributed, hybrid, multiple–ontology clinical–guideline library, and automated guideline–support tools. *Journal of Medical Bioinformatics,* 37:325–344.
13. P. Terenziani, S. Montani, A. Bottrighi, M. Torchio, G. Molino, G. Correndo (2005). Managing Clinical Guidelines Contextualization in the GLARE System. *AI*IA* 2005: 454–465.
14. E. Triantafyllou, E. Kokkinou, P. de Clercq, N. Peek, H. M. Korsten, A. Hasman (2005). Representation and execution of temporal criteria for guideline–based medical decision support at the Intensive Care Unit. *BNAIC* 2005: 239–246.
15. M. Vilain, H. Kautz (1986). Constraint propagation algorithms for temporal reasoning. In *Proceedings of the National Conference on Artificial Intelligence* (AAAI–86), 377–382.
16. World Wide Web Consortium (W3C) (2004). *OWL Web Ontology Language Overview.* W3C Recommendation 10 February 2004. D. L. McGuinness, F. van Harmelen eds.

Sistema experto para soporte diagnóstico en el postoperatorio de transposición de grandes arterias

Víctor Raúl Castillo[1], Xiomara Patricia Blanco Valencia[2], Álvaro Eduardo Durán[1], Gregorio José Mauricio Rincón Blanco[3] y Andrés Felipe Villamizar Vecino[3]

[1] MD, Fundación Cardiovascular de Colombia
inv_pediatria@fcv.org
[2] Ing. de Sistemas, Fundación Cardiovascular de Colombia
xblanco@fcvsoft.com
[3] Estudiantes Universidad Industrial de Santander

Resumen. En los mejores centros del mundo la mortalidad acumulada de cirugía cardiovascular pediátrica es cercana al 2%, en la Fundación Cardiovascular de Colombia, la incidencia acumulada de muerte en los últimos 8 años es del 5.8%. El gran volumen de información, la síntesis e interpretación se suele hacer manualmente en la mayoría de las Unidades de Cuidado Intensivo, haciéndolas un lugar susceptible al error médico y sitio ideal para la aplicación de una serie de herramientas informáticas, que permitan una mayor eficiencia en el cuidado de los pacientes. Uno de los campos de aplicación de la Inteligencia Artificial son los sistemas expertos, que solucionan problemas utilizando conocimientos basados en hechos y capacidad de razonamiento. Este tipo de tecnologías, realiza diferentes aportes a la práctica clínica: so-porte confiable para la toma de decisiones mostrando sugerencias, estandarizando actitudes, métodos y tratamientos, respetando la autonomía del profesional de la medicina.

1 Planteamiento de un problema clínico modelo

La Transposición de Grandes Arterias (TGA) se refiere al origen invertido de las grandes arterias del corazón. Hoy en día es posible corregir esta anormalidad en niños de 15 días de vida con un porcentaje de éxito que en los mejores centros del mundo es cercano al 100% [1]. En nuestro medio debido a las dificultades del recurso humano y en algunos casos técnicos el porcentaje de sobrevida puede ser inferior al 40%.

La mortalidad de esta patología en las unidades de cuidado intensivo (UCIs) es un indicador de calidad, debido a la alta complejidad del manejo postoperatorio, a la baja tolerancia al error, requiriéndose una estructura organizacional sofisticada, coordinación de esfuerzos de múltiples personas trabajando en equipo y altos niveles de conocimiento y desempeño técnico [1]. Se ha descrito que bajos índices de mortalidad asociados a esta patología, se asocian con mejores resultados en otros postoperatorios. Al comparar las UCIs de hospitales como los de Suecia, los cuales cuentan con los mismos recursos técnicos que las de los hospitales Colombianos, se encuentra que sus

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 146-151, 2006.

índices de mortalidad son menores. Esto último se ha atribuido al alto nivel de entrenamiento de su personal, con lo cual han logrado reducir el error humano.

2 Sistemas expertos

La disciplina conocida como Inteligencia Artificial (IA) ha generado en las últimas décadas un gran número de sistemas con supuestos rasgos de inteligencia[1] en los que interactúan unos datos de entrada con los de una base de información que contiene el conocimiento de un área específica. Un tipo bien conocido de este tipo de sistemas es el de los sistemas expertos (SE).

Una de las aplicaciones más importantes de los SE ha tenido lugar en el campo médico, donde éstos han sido utilizados para recoger, organizar, almacenar, poner al día y recuperar información médica de una forma eficiente y rápida, permitiendo aprender de la experiencia [2].

Dentro de los trabajos relacionados con el tema se pueden citar: INTERNIST-1/CADUCEUS, sistema de fácil uso sobre medicina interna. MYCIN, construido en Stanford que diagnostica enfermedades infecciosas de la sangre y receta los antibióticos apropiados y PUFF que diagnostica enfermedades pulmonares.

En la Universidad de Edimburgo se esta realizando una tesis de doctorado para evaluar condiciones de monitoreo en neonatos en una UCI. Dicho sistema detecta falsas alarmas causadas por diferentes factores y puede inferir el comportamiento de una variable [3].

Desde hace 2 años, la Fundación Cardiovascular de Colombia (FCV) ha ido desarrollando una herramienta de ayuda diagnóstica y terapéutica orientada a la reducción del error humano y a la mejora de procesos de atención en salud gracias a la toma oportuna de conductas para preservar la estabilidad clínica y predecir la complicación de niños atendidos en la UCI postoperatoria cardiovascular pediátrica. Dicha herramienta, podría en un futuro ser extendida a otras disciplinas y centros de atención de alta complejidad, con enormes beneficios tanto en el sector público como en el privado.

3 Aplicación del proyecto al modelo clínico de interés

La TGA uno de los diagnósticos congénitos más complejos por su manejo postoperatorio. Si adicionalmente se considera la gran cantidad de variables que se requiere integrar al realizar el ejercicio médico de su atención; encontramos explicación a

[1] En el área de la IA, el término inteligencia se utiliza de manera débil, haciendo referencia a cualquier comportamiento de apariencia inteligente, que puede ir del reconocimiento de un patrón perceptual determinado a la toma de decisiones basada en reglas preconstruidas. La mayoría de los sistemas no presupone una cierta teoría global de la inteligencia humana, ni pretende simular la función del cerebro, más bien, se limita a aprovechar técnicas algorítmicas existentes para resolver una situación concreta.

situaciones cotidianas en servicios asistenciales como las UCIs donde son comunes los errores de comunicación, la perdida e incorrecta trascripción de datos, y la demora en la obtención de registros que alerten sobre anormalidades y permitan anticiparse a posibles complicaciones en los pacientes [1]. Todo esto hace estos servicios particularmente susceptibles al error médico y el ambiente ideal para la aplicación de una serie de herramientas informáticas tradicionales o basadas en técnicas de IA, que permitan una mayor eficiencia en el cuidado de los pacientes.

4 Metodología y descripción de las etapas del proyecto

El desarrollo del sistema se planeó realizar por ciclos que constituyeran cada uno, una versión del producto. Cada ciclo se dividió en cuatro fases (inicio, elaboración, construcción y transición) que a su vez se dividieron en iteraciones que se planearon desarrollar a lo largo de 5 flujos de trabajo fundamentales: requisitos, análisis, diseño, implementación y pruebas.

Fase de elaboración:

En esta fase la labor primordial fue la selección de una arquitectura estable con el fin de planificar adecuadamente las labores de construcción. Se definieron aspectos como motor de inferencia, lenguaje de programación y plataforma.

Motor de inferencia: se analizaron catorce diferentes shells de motores de inferencia en aspectos como plataforma, lenguaje, estrategia de búsqueda, forma en que elige el conocimiento y posibilidad de incorporar metaconocimiento.

Al finalizar nuestro análisis se escogieron Jess y Clips como posibles shells[2] ya que cumplían con características como trabajo en plataformas Linux y Windows, integración con java[3] como lenguaje de programación, sistema de producción por encadenamiento hacia delante (de las mejores estrategias encontradas para diagnóstico médico)[4] y tenían implementado Rete como algoritmo de búsqueda.

Algoritmo Rete: del tiempo de ejecución el 90% se consume en el proceso de emparejamiento. El algoritmo Rete se basa en dos observaciones (suposiciones):

La memoria de trabajo es muy grande y cambia poco entre cada ciclo. El esfuerzo de emparejamiento depende de la razón de cambio de la memoria de trabajo en lugar del tamaño de ésta. Las condiciones de muchas reglas son similares. Rete procesa (compila) las reglas antes de ser usadas, localizando condiciones comunes y eliminando todas menos una [5].

Fue diseñado para facilitar el análisis temporal estático de los programas que lo usan, así propiedades como la no duplicación, la novedad, la especificidad y la prioridad de operación, se preservan. [6][7]

---

[2] El intérprete de comandos usado para interactuar con el núcleo de un sistema operativo.

[3] Java es un lenguaje de programación libre, independiente de la plataforma, muy extendido y con mucha importancia en el ámbito de Internet. [15]

Fase de construcción:

Se construyó el primer prototipo de TGA.

Construcción de la base de conocimiento: se diseñaron e implementaron las reglas del protocolo de manejo postoperatorio de TGA que incluyen aspectos como:

- Alarmas generadas por alteración de una variable.
  El especialista tiene la posibilidad de mirar la alteración de una variable en especial y el sistema le muestra un listado de los posibles diagnósticos asociados con la alteración de esta variable, teniendo en cuenta prioridades en las complicaciones posibles. Cuando se escoge uno de la lista, el sistema muestra un paralelo entre los signos asociados con los signos y síntomas que presenta el paciente para que el especialista pueda decidir si acepta o no el diagnostico. De esta manera el especialista va construyendo reglas en la práctica clínica y se esta generando un histórico para alimentar el sistema de aprendizaje a realizar en una etapa posterior.
- Otro tipo de reglas muestran consideraciones a tener en cuenta por posibles complicaciones dependiendo de la edad y peso del paciente.
- Un conjunto de reglas sugieren tratamientos a seguir dependiendo del tiempo del postoperatorio, edad, peso y signos del paciente.

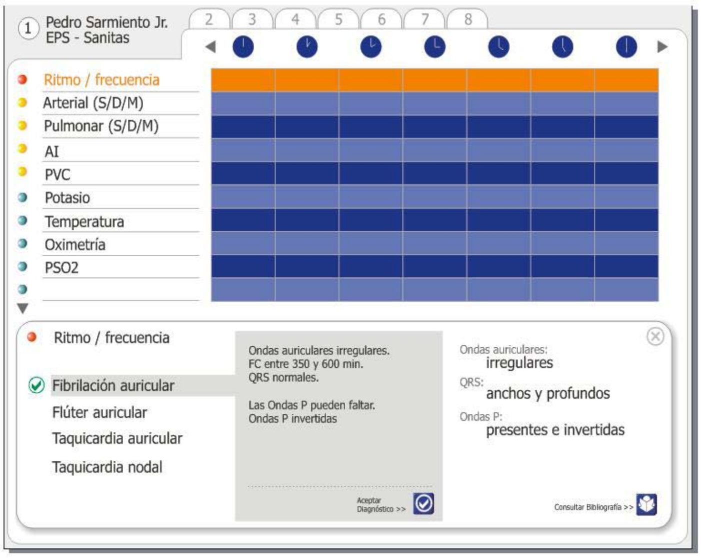


Fig. 1. Interfaz donde se pueden ver las alertas generadas por las variables, los diagnósticos asociados y los datos que el paciente presenta.

Hasta el momento se han realizado pruebas experimentales en dos niños en este tipo de postoperatorio obteniendo resultados favorables y observando que estas mismas reglas se pueden generalizar a casi todos los postoperatorios cardiovasculares de la UCI.

El sistema esta en capacidad de conectarse al monitor de signos vitales diseñado por el grupo de bioingeniería de la FCV [8], ventaja que le permite al SE disparar alarmas en tiempo real. Adicionalmente esta conectado a la base de datos de la HCE[4] de donde obtienen los valores de los signos no monitoreados en tiempo real.

Vale la pena resaltar que el diseño de este prototipo se dio gracias a un trabajo interdisciplinario de médicos e ingenieros, que inicialmente acordaron las características principales del sistema y visualizaron la necesidad de realizar un estudio de usabilidad[5] dada la complejidad de los sitios donde va a ser empleado.

Se considera que las fases de inicio, elaboración y construcción deben repetirse hasta que exista un sistema lo suficientemente robusto para volver a realizar pruebas. El paso siguiente fue el desarrollo de una interfaz de usuario para el ingreso de reglas y visualización de resultados teniendo en cuenta el estudio de usabilidad. Paralelamente a este proceso se probaran las reglas implementadas para el manejo de pacientes operados de TGA.

Sistema de aprendizaje: será desarrollado bajo el marco teórico que existe acerca de las Máquinas de Soporte Vectorial (MSV). Los progresos importantes en teoría de aprendizaje basados en estadística han introducido nuevos paradigmas tales como MSV cuya principal característica es la estructura de un algoritmo de aprendizaje el cual consiste en la solución de un simple problema cuadrático. [9]. Las MSV se pueden ver también como un nuevo método para el entrenamiento de modelos polinómicos, redes neuronales, modelos borrosos, entre otros [9].

5 Discusión de los resultados preliminares

Al revisar la historia de los sistemas basados en el conocimiento, la mayoría de ellos son extremadamente pequeños en cuanto al volumen de sus bases de conocimiento y base de hechos. Es decir, el universo de estos sistemas es simple y limitado, y en la mayoría de los casos se centran exclusivamente en resolver una situación o un problema determinado. Partiendo del reconocimiento de esa situación de estrechez en el ámbito de los sistemas basados en el conocimiento, el presente proyecto pretende construir una base de conocimiento de un volumen radicalmente mayor en diagnóstico y tratamiento de enfermedades cardiovasculares. Adicionalmente puede llegar a tener un gran impacto clínico si se asocia con una disminución significativa de indicadores como la mortalidad en postoperatorios cardiovasculares. La complejidad de plasmar el conocimiento médico requiere de mucho tiempo y dedicación, el presente proyecto le permite al especialista construir las reglas durante la práctica clínica. El presente proyecto, podría en un futuro ser extendido a otras disciplinas y centros de atención de alta complejidad, lo cual tendría un gran impacto en el campo asistencial.

---

[4] Sistema software de Historia Clínica Electrónica.

[5] **Usabilidad** es la efectividad, eficiencia y satisfacción con la que un producto permite alcanzar objetivos específicos a usuarios específicos en un contexto de uso específico" ISO/IEC 9241.

Referencias

[1] Leval Marc, MD. Carthey Jane, PhD. Wright David, PhD. And all United Kingdom pediatric cardiac centers. Human Factors and Cardiac Surgery: a multicenter study. Toracic Cardiovasc. Surg. -01-APR-2000;119(4 pt) 661-72 from NIV MEDLINE.

[2] E. Castillo, J.M. Gutiérrez, and A.S. Hadi. Expert Systems and Probabilistic Network Models. Springer Verlag, New York, 1997. 600 pages.ISBN: 0- 387-94858-9.

[3] Williams Christopher K. I., Quinn John, McIntosh Neil. Factorial Switching Kalman Filters for Condition Monitoring in Neonatal Intensive Care. November 2005.

[4] Montagut Martha Vitalia. Principios de inteligencia artificial y sistemas expertos. Ediciones UIS 2000.

[5] L. Martin, W. Taylor, S. Meadows and K. Freeman. CLIPS Application Abstracts. November 1st 1997.

[6] Third Conference on CLIPS. September 12-14, 1994. Lyndon B. Johnson Space Center. 19 – 22.

[7] CLIPS Version 5.1 User's Guide, NASA Lyndon B. Johnson Space Center, Software. Technology Branch, Houston, TX, 1991.

[8] Proyecto aprobado por COLCIENCIAS. “Diseño y construcción de un prototipo para monitoreo presencial y remoto de signos vitales de pacientes en estado crítico”. Contrato No. 332-2004, Código No. 6566-14-172117. Fundación Cardiovascular de Colombia.

[9] Vladimir N. Vapnik, The Nature of Statistical Learning Theory. Springer, 1998.

Decisión multi-atributo basada en órdenes de magnitud

Núria Agell[1], Mónica Sánchez[2], Francesc Prats[2] y Xari Rovira[1]

[1] ESADE Universitat Ramon Llull, Av. Pedralbes 62. 08034 Barcelona
{Nuria.Agell, Xari.Rovira}@esade.edu
[2] Universitat Politècnica de Catalunya, Dept MA2, Jordi Girona, 1-3
08034 Barcelona
{Monica.Sanchez, Francesc.Prats}@upc.edu

Resumen. En este artículo se propone un método de síntesis de información cualitativa en órdenes de magnitud para la evaluación y ayuda a la toma de decisiones multi-atributo. Éste permite la elección de una entre diversas alternativas, caracterizadas por variables cualitativas con valores en un espacio de órdenes de magnitud absolutos. El método consiste en representar las diferentes alternativas por medio de etiquetas k-dimensionales, interpretándose cada una de ellas como la conjunción de los k intervalos correspondientes a las etiquetas cualitativas de las variables de entrada. Se da un método para escoger la mejor alternativa, basado en la comparación de distancias a una etiqueta de referencia previamente construida. Se introduce una distancia en el conjunto de las etiquetas k-dimensionales, que define un orden total en dicho conjunto a partir de la etiqueta de referencia, se propone un método de elección basado en este orden, y se demuestra la consistencia del método.

1 Introducción

En procesos de decisión multi-atributo, la evaluación de alternativas depende de ciertos factores o variables de entrada [4], [9] que, en ocasiones, son difíciles de valorar de forma exacta. El método de ayuda a la toma de decisiones que se presenta en este trabajo es especialmente adecuado cuando el objetivo es realizar una evaluación a partir de variables expresadas en órdenes de magnitud. Estas descripciones cualitativas se consideran cuando los valores numéricos no se conocen con precisión, o bien cuando los órdenes de magnitud y las tendencias de las variables son más relevantes que sus valores numéricos exactos. También aparecen en problemas en los que las variables están medidas sobre escalas ordinales.

Este trabajo supone la adaptación al caso de variables definidas sobre espacios de órdenes de magnitud absolutos de un trabajo previo basado en álgebra intervalar [6]. A diferencia de las álgebras intervalares los espacios de órdenes de magnitud parten de un conjunto predeterminado de etiquetas que es fijado a priori y no depende de los extremos de los intervalos. Como hipótesis de trabajo se considera que el valor asignado a las alternativas es una función creciente respecto a las variables de entrada, es decir, a mayores valores de cada variable

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 152-160, 2006.

de entrada corresponde mayor valor de la alternativa. En el caso de tener dependencia decreciente con respecto de alguna variable, se reemplazará dicha variable por su opuesta, cambiando el signo de sus valores.

El método que se propone consiste en sintetizar la información inicial via un rectángulo de $\mathbb{R}^k$ correspondiente a una etiqueta cualitativa k-dimensional y realizar la evaluación utilizando una distancia a una etiqueta de referencia. Está basado en una generalización cualitativa de los métodos de "goal programming" llamados "métodos de puntos de referencia" para optimización vectorial y ayuda a la decisión [3], [7]. En general, los métodos de puntos de referencia para la optimización en $\mathbb{R}^n$ eligen como alternativa óptima los puntos que están a menor distancia de un punto de referencia previamente fijado en el espacio (el "goal" o la meta que se desea alcanzar) [2]. En este trabajo, la optimización en el conjunto de etiquetas k-dimensionales se realiza escogiendo una etiqueta de referencia "realista" para el problema a solucionar y que no se fija a priori: la etiqueta de referencia propuesta es el supremo según el orden natural del conjunto de las alternativas disponibles. Se garantiza así la consistencia con el caso en que a priori ya se tiene una alternativa mejor que todas las demás.

La metodología propuesta es de interés en áreas muy diversas. En concreto se pueden considerar aplicaciones tanto en temas de evaluación de candidatos (estudiantes en procesos de aprendizaje, aspirantes en procesos de selección de personal,...) como en la gestión de proyectos (arquitectónicos, de ingeniería civil, empresariales,...) [5] , o en la toma de decisiones en áreas como finanzas ([1]) y marketing.

En la sección 2 se presentan los modelos cualitativos de órdenes de magnitud absolutos que constituyen el marco de referencia de este trabajo. En la sección 3 se propone una representación cualitativa de las alternativas en el conjunto parcialmente ordenado $\mathcal{E}$ de las etiquetas cualitativas k-dimensionales de órdenes de magnitud absolutos. En la cuarta sección se define una distancia en el conjunto $\mathcal{E}$. En la sección 5 se define un orden total en $\mathcal{E}$, de tal manera que se podrá establecer un ranking en el conjunto dado de alternativas. Se establece la propiedad de consistencia del método de elección, que determina la etiqueta de referencia en el caso que ya previamente exista una alternativa mejor que todas las demás, y se generaliza al caso de etiquetas sin máximo, escogiendo la etiqueta de referencia como el supremo de estas etiquetas k-dimensionales respecto del orden parcial natural en $\mathcal{E}$. Finalmente se presentan las conclusiones obtenidas y se plantean problemas abiertos.

2 Modelos cualitativos de órdenes de magnitud absolutos

En el modelo absoluto de órdenes de magnitud ([8]) se trabaja con un número finito de etiquetas cualitativas obtenidas via una discretización de la recta real.

El modelo utilizado es una generalización del modelo introducido en [8]. El número de etiquetas que se escogen para describir la realidad no es fijo y dependerá de las características de la variable representada. Se construye el modelo de órdenes de magnitud absolutos via una partición de un intervalo de la recta real

$[a_1, a_{n+1}]$ construida a partir de un conjunto de puntos frontera $\{a_1, \ldots, a_{n+1}\}$ (no necesariamente simétrica, ni necesariamente con el mismo número de etiquetas positivas que negativas):

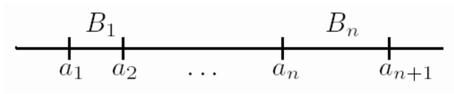


Fig. 1. La discretización

Cada clase de la partición es una descripción básica y se representa por una etiqueta en el conjunto S^*:

$$S^* = \{B_1, ..., B_n\}, \text{con } B_i = [a_i, a_{i+1}],\ i = 1, \ldots, n$$

El llamado espacio completo de cantidades S queda definido por extensión de S^* como $S = S^* \cup \{[B_i, B_j]/B_i, B_j \in S^*,\ \text{con } i < j\}$, siendo $[B_i, B_j]$ la etiqueta correspondiente al mínimo intervalo cerrado de la recta real que contiene a B_i y a B_j. Es decir, si $B_i = [a_i, a_{i+1}], B_j = [a_j, a_{j+1}]$, entonces $[B_i, B_j] = [a_i, a_{j+1}]$. [8].

La relación $\leq_P$, *ser más preciso* que (dados $E, E' \in S$, E es más preciso que E', $E \leq_P E'$, si $E \subseteq E'$) es una relación de orden parcial en S. Para cualquier $E \in S$, la base de E es el conjunto: $B_E = \{B \in S^* | B \leq_P E\}$ y, dado un elemento básico $B \in S^*$, la B-expansión de E, E_B, es la mínima etiqueta de S que es menos precisa que E y que B, i.e., el menor intervalo respecto de la inclusión que contiene a E y a B. Nótese que E_B no depende de los valores de los puntos frontera que determinan la partición de la recta real.

Es importante remarcar que, en el problema que se plantea en este artículo, las variables pueden estar definidas en espacios de diferente granularidad (diferente número de etiquetas básicas) y, evidentemente, cada una tendrá su propia discretización.

3 Representación de las alternativas: el conjunto parcialmente ordenado $\mathcal{E}$

En el problema de decisión multi-atributo que se plantea, cada una de las alternativas está caracterizada por los valores de k atributos o variables de entrada, y estos valores vienen dados por etiquetas cualitativas pertenecientes cada una a un espacio de órdenes de magnitud. Sea S_i el espacio de órdenes de magnitud asociado a la variable i-ésima, cuyo conjunto de elementos básicos será ${S_i}^*$, tal como se ha introducido en el apartado anterior. Se define el conjunto de posibles alternativas $\mathcal{E}$ como:

$$\mathcal{E} = S_1 \times \ldots \times S_k = \{(E_1, \ldots, E_k)\ |\ E_i \in S_i\ \forall i = 1, \ldots k\}$$

De esta forma, cada alternativa se puede interpretar como una etiqueta k-dimensional, i.e. un rectángulo k-dimensional cuyas componentes son etiquetas cualitativas.

Con el objetivo de comparar alternativas, se define una relación de orden parcial en cada componente S_i inducida por el orden intervalar natural en S_i^* ($B_i \leq B_j \Longleftrightarrow x \leq y \,\forall\, x \in B_i, \forall\, y \in B_j$):

Sean $E = [B_i, B_j]$ y $E' = [B_i', B_j']$ dos etiquetas de un espacio de órdenes de magnitud S, con $B_i, B_j, B_i', B_j' \in S^*$:

$$E \leq E' \Longleftrightarrow B_i \leq B_i' \,\text{y}\, B_j \leq B_j'$$

Esta relación es una relación de orden parcial en S. De ella se induce, por extensión al producto cartesiano, la siguiente relación de orden parcial en $\mathcal{E}$:

$$(E_1, \ldots, E_k) \leq (E_1', \ldots, E_k') \Longleftrightarrow E_i \leq E_i', \;\; \forall i = 1, \ldots, k.$$

Éste es un orden parcial ya que evidentemente no todo par de etiquetas k-dimensionales son comparables (ver figura 2):

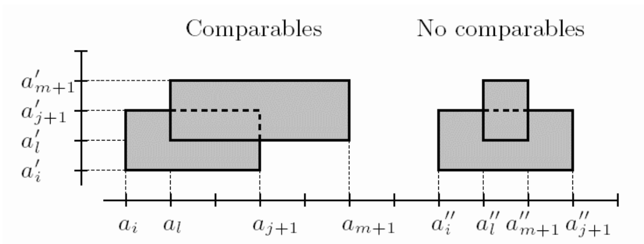


Fig. 2. El orden parcial $\leq$ en $\mathcal{E}$

Por la consideración hecha en la sección 1 (mayores valores de cada variable significan mayor valor de la alternativa), se tiene que $E \leq E'$ corresponde a que E' es mejor alternativa que E.

En la siguiente sección se propone una distancia entre alternativas que permitirá calcular distancias de una alternativa cualquiera a la de referencia, tal como se realiza habitualmente en los métodos de "goal programming", y de esta manera establecer después el orden total.

4 Distancia en el conjunto $\mathcal{E}$

La definición de una distancia entre etiquetas k-dimensionales está basada en una función de localización que permitirá medir la posición relativa entre alternativas utilizando una distancia en $\mathbb{R}^{2k}$. Cada elemento E en S se codifica por un par $(l_1(E), l_2(E))$ de números enteros: $l_1(E)$ es el número de elementos básicos en S^*

que están "entre" B_1 y la base de E, y $l_2(E)$ es el número de elementos básicos en S^* que están "entre" la base de E y B_n. Este par de números permiten "localizar" todos los elementos de S.

La definición formal de la función de localización es $l : S \to Z^2$ tal que:

$$l(E) = (l_1(E), l_2(E)) = (-\mathrm{card}(B_{E_{B_1}}) + \mathrm{card}(B_E), \mathrm{card}(B_{E_{B_n}}) - \mathrm{card}(B_E))$$

La función de localización codifica etiquetas de S por medio de puntos del plano euclídeo, de manera que la distancia euclídea entre ellos permitirá definir una distancia entre etiquetas. Esta función de localización se puede extender a cualquier alternativa definida por k variables de órdenes de magnitud. La extensión de l a $S_1 \times \ldots \times S_k$ es la función $L : S_1 \times \ldots \times S_k \to Z^{2k}$ definida por:

$$L(E_1, \ldots, E_k) = (l(E_1), \ldots, l(E_k)) = (l_1(E_1), l_2(E_1), \ldots, l_1(E_k), l_2(E_k)).$$

Esta función de localización permite medir la similitud entre alternativas. Una vez se han codificado las etiquetas k-dimensionales por medio de $2k$-tuplas de números enteros, se está en condiciones de definir una distancia entre etiquetas k-dimensionales.

Se define:

$$d : \mathcal{E}^2 = (S_1 \times \ldots \times S_k)^2 \to [0, +\infty)$$

$$(E, E') \mapsto \sqrt{(L(E) - L(E')) \cdot R \cdot (L(E) - L(E'))^t}$$

donde R representa cualquier métrica en $\mathbb{R}^{2k}$.

Es decir, si $E = (E_1, \ldots, E_k)$ y $E' = (E'_1, \ldots, E'_k)$

$$d(E, E') =$$
$$\sqrt{(l_1(E_1) - l_1(E'_1), \ldots, l_2(E_k) - l_2(E'_k)) \cdot R \cdot (l_1(E_1) - l_1(E'_1), \ldots, l_2(E_k) - l_2(E'_k))^t}$$

Esta función d cumple las propiedades de distancia en $\mathbb{R}^{2k}$ por provenir del producto escalar dado por R.

Cabe destacar que esta definición permite, escogiendo una métrica R adecuada, trabajar en el caso eventual en que las variables consideradas requieran distinta ponderación.

5 Elección de la mejor alternativa

En el siguiente subapartado se define un orden total $\trianglelefteq$ en $\mathcal{E}$ a partir de una distancia en $\mathcal{E}$ y de una etiqueta k-dimensional de referencia $\overline{E}$, de tal manera que el conjunto de etiquetas $E_1, \ldots, E_n$ correspondientes a las alternativas dadas podrá escribirse en forma de cadena $E_{i_1} \trianglelefteq \cdots \trianglelefteq E_{i_n}$. La alternativa E_{i_n} correspondiente al máximo de la cadena será la escogida como la mejor de entre las dadas.

5.1 Un orden total en $\mathcal{E}$

Sea $\overline{E} \in \mathcal{E}$ una etiqueta k-dimensional cualquiera, a la que llamaremos etiqueta de referencia.

Si designamos por d a la distancia definida en $\mathcal{E}$ de la sección anterior, la relación en $\mathcal{E}$:

$$E \preceq E' \Longleftrightarrow d(E', \overline{E}) \leq d(E, \overline{E})$$

es un preorden, es decir, reflexiva y transitiva.

Este preorden induce una relación de equivalencia en $\mathcal{E}$ mediante:

$$E \equiv E' \Longleftrightarrow E \preceq E' \ , \ E' \preceq E \Longleftrightarrow d(E', \overline{E}) = d(E, \overline{E}).$$

En el conjunto cociente $\mathcal{E}/\equiv$ la relación entre clases:

$$\text{clase}\,(E) \trianglelefteq \text{clase}\,(E') \Longleftrightarrow E \preceq E' \Longleftrightarrow d(E', \overline{E}) \leq d(E, \overline{E})$$

ya es de orden, y evidentemente, total.

De esta manera, dado un conjunto de alternativas $E_1, \ldots, E_n$, pueden ordenarse como una cadena respecto de su proximidad a la etiqueta de referencia: $\text{clase}\,(E_{i_1}) \trianglelefteq \cdots \trianglelefteq \text{clase}\,(E_{i_n})$. Las alternativas de una misma clase, es decir, las que están todas a la misma distancia de $\overline{E}$, se considerarán igual de buenas, por lo que se hará de ahora en adelante un abuso de notación cambiando $\preceq$ por $\trianglelefteq$: $E_{i_1} \trianglelefteq \cdots \trianglelefteq E_{i_n}$.

5.2 Consistencia del método de elección

El método de elección de la mejor alternativa, via distancias a una etiqueta de referencia previamente fijada, tiene sentido cuando a priori ninguna de ellas es mejor que todas las demás, es decir, cuando el conjunto $\{E_1, \ldots, E_n\}$ no tiene máximo según el orden $\leq$.

Ahora bien, en el caso de que $\{E_1, \ldots, E_n\}$ tenga un máximo E_m según $\leq$, el método de elección propuesto será consistente si proporciona como mejor alternativa la misma E_m.

Formalmente, la etiqueta de referencia $\overline{E}$ dará lugar a un método consistente cuando se cumpla que, dada una colección cualquiera $E_1, \ldots, E_n \in \mathcal{E}$, se tiene:

$$\exists m \in \{1, \ldots, n\} \ \ E_i \leq E_m \ \forall\, i = 1, \ldots, n \Longrightarrow E_i \trianglelefteq E_m \ \forall i = 1, \ldots, n.$$

Si para cualquier colección de etiquetas k-dimensionales $E_1, \ldots, E_n$ con máximo E_m se toma como etiqueta de referencia $\overline{E} = E_m$, entonces el método es consistente, ya que $d(\overline{E}, \overline{E}) = 0$.

5.3 Elección de la etiqueta de referencia

Se ha visto que la consistencia del método de elección propuesto queda asegurada, en el caso de una colección de etiquetas con máximo, si se toma como etiqueta de referencia dicho máximo. La generalización natural al caso de una colección de etiquetas cualesquiera es la siguiente:

Dados $E_1, \ldots, E_n$ cualesquiera, se tomará como etiqueta de referencia el supremo de estas etiquetas respecto del orden parcial $\leq$ (ver figura 3):

$$\overline{E} = \sup\{E_1, \ldots, E_n\}.$$

Es decir, si $E_r = (E_1^r, \ldots, E_k^r)$, con $E_i^r = [B_{i_1}^r, B_{i_2}^r]$ para todo $i = 1, \ldots, k$, y para todo $r = 1, \ldots, n$, entonces $\overline{E} = (\overline{E}_1, \ldots, \overline{E}_k)$, donde

$$\overline{E}_i = [\max\{B_{i_1}^1, \ldots, B_{i_1}^n\}, \max\{B_{i_2}^1, \ldots, B_{i_2}^n\}].$$

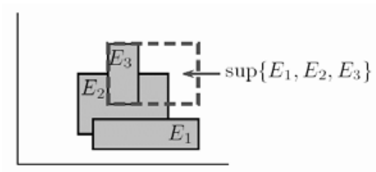


Fig. 3. Etiqueta de referencia: el supremo

Nótese que en el caso particular de etiquetas con máximo, este supremo es precisamente este máximo, por lo que esta elección de la etiqueta de referencia hace que se mantenga la propiedad de consistencia del método.

5.4 Elección de la mejor alternativa

Finalmente, se resumen los pasos del método propuesto para la elección de la mejor alternativa:

1. Definir una distancia d en $\mathcal{E}$ asociada a una métrica.
2. Calcular la etiqueta de referencia $\overline{E}$: el supremo del conjunto de alternativas.
3. Asignar a cada etiqueta E el valor $d(E, \overline{E})$ y así obtener un ranking en el conjunto de alternativas.
4. Elegir como la mejor alternativa el máximo de la cadena, es decir, la (o las) de distancia mínima.

6 Conclusión

En este trabajo se propone un método para la evaluación de alternativas multi-atributo basado en la utilización de distancias a un elemento de referencia. La utilización de variables valoradas en espacios de órdenes de magnitud absolutos permite abordar los problemas de aplicación con información cualitativa de forma consistente.

La metodología presentada permite, por un lado, tratar los conceptos cualitativos que algunas aplicaciones conllevan, y por otro, generalizar los métodos de "goal programming" sin necesidad de conocer previamente el objetivo ideal.

Es importante destacar que, sin obviar el carácter intervalar de las etiquetas cualitativas, en la propuesta que se presenta en este artículo no es necesario conocer los valores de las fronteras de las discretizaciones consideradas.

En el marco del proyecto de investigación AURA ("sistemas de aprendizaje AUtomático con RAzonamiento cualitativo: herramientas inteligentes de apoyo a la decisión aplicadas a las finanzas y al marketing") financiado por el MCyT (TIN2005-08873-C02-01 y TIN2005-08873-C02-02), se pretende aplicar el método presentado a la evaluación de empresas según su nivel de riesgo de crédito. En concreto, con este fin, se toman variables relativas al tamaño de la empresa, a la actividad que desarrolla, a su financiaciún, liquidez, rentabilidad y volatilidad de las cotizaciones bursátiles. Asimismo se tienen en cuenta datos cualitativos de las empresas como son el país en que se localizan y el sector de actividad al que corresponden. La técnica descrita permite ordenar dichas empresas de mejor calidad de crédito a máximo riesgo. Una vez ordenadas se podrán comparar los resultados con las clasificaciones que proporcionan las agencias de rating. El objetivo final de esta aplicación es el de validar la técnica y, al mismo tiempo, encontrar un método que permita tomar decisiones de inversión frente a distintas alternativas.

Como trabajo futuro, se pretende analizar la eficacia de otras distancias cualitativas y la aplicabilidad de la metodología presentada a otras áreas de conocimiento.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el proyecto coordinado de investigación del MEC (Ministerio de Educación y Ciencia) AURA (TIN2005-08873-C02-01 y TIN2005-08873-C02-02).

References

1. Agell, N., Rovira, X., Ansotegui, C., Sánchez, M., Prats, F. "Predicting Financial Risk by Qualitative Reasoning Techniques". Actas 15th International Workshop on Statistical Modelling. 2000. Bilbao, España.
2. González Pachón, J., Romero López, C. Aggregation of partial ordinal rankings. An interval goal programming approach revista: Computers and operations research. 28 (2001) 827-834

3. Kallio,M., A. Lewandowski and W. Orchard-Hays An Implementation of the Reference Point Approach for Multi-Objective Optimization. WP-80-35, IIASA, Laxenburg (1980)
4. Keeney, R.L., and Raiffa, H. (1993). Decisions with multiple objectives preferences and value trade-offs, Cambridge University Press.
5. Ormazabal, G. (2002). "IDS: A new integrated decision system for construction project managemen". PhD Thesis, Department of Construction Engineering, Technical Univ.of Catalonia (UPC), Barcelona, España.
6. Prats,F., Sánchez,M., Agell,N., Omazabal,G. "Un método de evaluación con información imprecisa para la ayuda a la decisión multi-atributo". Actas XI Conferencia de la Asociación Española para la Inteligencia Artificial, CAEPIA 2005, Vol 1, pp. 393-402, Santiago de Compostela, España.
7. Romero López, C., Tamiz, M., Jones, D. Comments on goal programming, compromise programming and reference point method formulations: linkages and utility interpretations-A reply revista: Journal of the Operational Research Society 52 (2001) 962-965.
8. Travé-Massuyès,L. and Dague,P. "Modèles et raisonnements qualitatifs". Hermès Science, 2003, Paris, France.
9. Wierzbicki, A.P. The use of reference objectives in multiobjective optimization. In G. Fandel, T. Gal (eds.): Multiple Criteria Decision Making; Theory and Applications, Lecture Notes in Economic and Mathematical Systems. 177 Springer-Verlag, Berlin-Heidelberg (1980) 468-486

Determinismo, autoconfiguración y posibilidades alternativas en la filosofía de la mente y de la acción de Daniel C. Dennett

Juan José Colomina Almiñana y Vicente Raga Rosaleny

Departament de Metafísica y Teoria del Coneixement, Universitat de València
Vicente.Raga@uv.es

Resumen. La intención del presente artículo es, en un primer momento, mostrar las virtudes y las aporías de un intento de dar sentido a la idea de autoconfiguración o autoformación en un mundo determinista. El intento en cuestión es el que representa Daniel C. Dennett, cuya propuesta analizaremos partiendo de la crítica escéptica que Galen Strawson presenta a toda posible idea de autoformación. En segunda instancia, pretendemos analizar el tratamiento que Dennett hace del Principio de Posibilidades Alternativas (PPA), así como las posibles consecuencias que suponen concebir la sobredeterminación de la acción humana y que, consideramos, Dennett pasa por alto.

El supuesto central que hace tambalear nuestra presuposición de que somos seres libres y moralmente responsables de nuestras acciones es que nuestro modo de actuar depende radicalmente de nuestro modo de ser, y que en la medida que no somos responsables de nuestro modo de ser, no podemos serlo de nuestro modo de actuar. Aunque nuestro modo de actuar pueda modificar nuestra manera de ser, siempre habrá un momento anterior, un modo de ser previo, una naturaleza o instancia anterior que nos haya permitido actuar de ese modo.

Un intento de superar esta dificultad es el de Robert Kane, con su idea de acciones auto-formativas (*self-forming actions*), unas acciones que han de permitir la exigencia de control último, es decir, ser responsables, controlar nosotros mismos nuestra manera de ser y no sólo nuestras acciones. Estas acciones auto-formativas supondrían una especie de hiato en la cadena causal, un instante de retroactividad de la acción sobre el modo de ser, una retroacción crucial sobre la que el sujeto ha de tener un control total. Aquí, en el caso de las acciones auto-formativas del yo (*self*) se tiene que producir un hueco en el determinismo causal, un momento de indeterminación que puede tener consecuencias para la racionalidad de la acción.

Este es quizá uno de los puntos débiles del argumento de Kane, pues ese hueco en la determinación causal, ese hiato, no permite salvar la dificultad de arbitrariedad, de irracionalidad incluso en la acción. Así pues, habría que indagar en la naturaleza de ese vacío. Quizá el problema no se solvente planteando ese vacío en la cadena causal, que es necesario para las acciones auto-formativas, sino atendiendo al tipo de causas que producen un tipo u otro de acción. Este primer punto sería el que un compatibilista trataría de hacer valer para modificar y corregir la posición de Kane: no se trata

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 161-173, 2006.

tanto de suspender la cadena causal en el interior de la facultad de razonamiento práctico como de apreciar los distintos tipos de causa que operan en ella, una tipología que permitiría salvar la responsabilidad moral. Sin embargo, un escéptico radical, como es el caso de Galen Strawson, no necesitaría recurrir a este argumento, sino que dirigiría su crítica a la línea de flotación tanto del compatibilista como del incompatibilista. En el caso de las acciones auto-formativas, Strawson apela a la paradoja de la *causa sui* ya denunciada por Nietzsche. El argumento de Strawson tiene el siguiente esquema: dado que mi elección de ser de una cierta manera depende de mi naturaleza, tengo que haber elegido (y ser responsable) de esa naturaleza mía, lo que supone que la elegí y la configuré responsablemente, pero siendo ya de una manera determinada (teniendo una naturaleza N –1). Aquí entramos en una regresión al infinito que invalidaría la argumentación de las acciones auto-formativas en un mundo determinista.
En lo que sigue vamos a intentar mostrar y analizar el intento de solución a este problema por parte de Daniel C. Dennett, especialmente en lo que concierne a la idea de auto-formación del yo (*self*).

La principal crítica que Dennett le dirige a Kane es que recurra a la idea de un vacío en la facultad de racionalidad práctica, un hueco que permita una acción auto-formativa indeterminada, lo que le lleva al problema de cómo hacer de una decisión indeterminada una decisión del sujeto, y no algo que le sucede (es decir, mantener el requisito de control en la indeterminación). Como buen compatibilista, Dennett quiere argumentar que las acciones auto-formativas, la formación del carácter «no necesita el indeterminismo que inspiró su creación» (Dennett, 2004: 149), sino que requiere cierta cadena causal evolutiva y determinista para ser posible.

De hecho, Dennett presenta un argumento muy peculiar para criticar las acciones auto-formativas de Kane (Dennett parece emparentar como veremos luego este argumento con el de Strawson). El argumento, falaz evidentemente, es el siguiente:

1. Todo mamífero tiene a un mamífero como madre.
2. En caso de que hayan existido los mamíferos, sólo ha podido ser en número finito.
3. Pero la existencia de un solo mamífero supone, en razón de 1, que debe haber existido un número infinito de mamíferos, lo que contradice 2. De la contradicción se sigue que no hay mamífero alguno.

Este juego argumentativo le sirve a Dennett para criticar a Kane la idea de las acciones auto-formativas (AA), pues se parecerían mucho al Mamífero primordial y fantástico que podría evitar el regreso al infinito. Además, Dennett apunta en una dirección interesante: ¿cómo discernir una acción auto-formativa de una que no lo es? ¿Cuántas acciones auto-formativas hacen falta para romper la regresión? Pero a su vez, desde nuestro punto de vista, ofrece un intento de superación del regreso al infinito al que nos enfrenta Strawson. Dennett encarna ese argumento regresivo de Strawson en la historia de la evolución, y al tratar de darle contenido histórico-biológico, éste parece adquirir un matiz distinto.

Así pues, Dennett indica ya su vía de estudio: insertar el problema de la libertad y la responsabilidad en la evolución biológica e histórica. Pero si tenemos en cuenta la argumentación de Dennett en sus diversas obras, no sólo se enfrenta a las AA de

Kane, sino también al tipo de argumento que presenta Strawson para llevarnos al escepticismo.

Para G. Strawson esta premisa es necesaria en su argumento, y en cierto modo Dennett impugna de partida ese supuesto que introdujimos al principio: que para ser responsables de nuestro modo de actuar hemos de ser responsables de nuestro modo de ser. Esta premisa de la propiedad transitiva de la responsabilidad entre el ser y la acción marca uno de los puntos críticos en el debate sobre la responsabilidad moral. En principio, el argumento escéptico ha de ser válido tanto para un mundo determinista como para uno indeterminista, y esto es un punto que hace dudar a Dennett de su validez, y defender que en un mundo determinista tal y como lo es el nuestro, marcado por la evolución, el argumento escéptico no tiene validez. Incluso teniendo en cuenta una versión indeterminista del argumento escéptico, en la que nuestro modo de ser se debería a la fortuna, Dennett trata de salvar la idea de auto-formación. Posteriormente abordaremos esta variante. Por el momento, tengamos en cuenta que, por una parte, Dennett no acepta el supuesto del argumento escéptico (Dennett, 1984: 84); pero que por otro, lo asume para salvar la dificultad de un yo construido por otra instancia, sin requerir que sea *causa sui*, pues se trata de una gradación de niveles de responsabilidad.

Sin embargo, llegados a este punto la carga de la prueba recae en aquél que rechaza la premisa: ¿cuál es la alternativa, dado que se acepta (con Strawson) que los sujetos no pueden ser *causa sui*? La inspiración de Dennett aquí es paradójicamente nietzscheana: del mismo modo que Strawson cita a Nietzsche para argumentar contra la *causa sui* que supondría el argumento de Kane, Dennett lo utiliza para argumentar a favor de una concepción de la creación de uno mismo, del propio yo (*self*), llegando a unas consecuencias que incluso el mismo Strawson parece aceptar, lo cual nos ha de poner en guardia y cuestionarnos si la noción de auto-formación del yo y la naturaleza del yo que defiende Dennett es lo suficientemente densa y resistente como para soportar el peso de la responsabilidad moral.

La creación del yo tiene dos desarrollos centrales en la teoría de Dennett, el primero en el contexto de la teoría de la evolución y el segundo en la construcción narrativa de cada individuo particular. Es decir, habría una vertiente filogenética (relativa a la especie) y otra ontogenética (relativa al individuo). Desde ambas perspectivas, el elemento clave es el lenguaje. De acuerdo con Dennett, en la evolución animal habría un momento en que la aparición del lenguaje introduce un nivel de auto-reflexión en los sujetos humanos, al punto que permite objetivar y distanciarse de las valoraciones, creencias y deseos que poseemos. El punto crucial es que para Dennett sólo podemos aceptar las acciones auto-formativas que plantea Kane en tanto que acciones narrativas, lingüísticas, en las que nosotros mismos tejemos nuestro pasado, nuestra memoria con nuestras acciones presentes (Dennett, 2004: 283). Aquí es cuando Dennett plantea su idea del yo o *self* como centro de gravedad narrativa. El desarrollo de esta idea se dedica más bien al nivel ontogenético, individual, pues da por demostrada naturalmente la cesura en la historia de la evolución con la aparición de un lenguaje que permite distinguir entre criaturas dignas de ser imputadas con responsabilidad moral y aquellas que no lo son. El lenguaje sería “ese algo especial” que nos permitiría «convertir las razones en objetos para la reflexión y el refinamiento» (Clark, 2002: 190). La historia evolutiva biológica se entrelaza con las consideraciones histórico-

políticas, puesto que Dennett incluye los contextos sociales en los que se fomenta y permite el uso del lenguaje en orden a dar razones, justificar y criticar las acciones como contextos que permiten un mayor desarrollo de agentes moralmente responsables.

Así pues, el nudo central de la argumentación de Dennett que queremos someter a análisis crítico es la aparición en escena del yo o *self* como peldaño que le permite "auparse a la libertad": en la construcción de este peldaño habrá que verificar la solidez de éste para no dar un paso en falso.

Desde la perspectiva filogenética, Dennett concedía una importancia central al lenguaje, como la piedra clave que sostiene el peso del arco y de la bóveda que permite un tratamiento especial de los seres humanos como agentes moralmente responsables. Aceptado este punto, y trasladando el lenguaje al nivel individual hemos de analizar en qué medida es condición de la responsabilidad moral.

Como hemos avanzado, la posición de Dennett parece apoyarse en la intuición nietzscheana de "llegar a ser el que eres", de construirse a uno mismo, de recrearse mediante el uso del lenguaje. Uno de los primeros rasgos que atribuye el autor a su definición de *self* es que es «un yo (*self*) es, sobre todo, un locus de auto-control (*self-control*)» (Dennett, 1984: 81). La condición de control es el rasgo definitorio del yo como agente digno de responsabilidad moral; en suspenso queda el grado de ultimidad que le concederá a esa instancia de control (Kane). De este modo, habrá que calibrar si la concepción del yo o *self* de Dennett puede dar cuenta de esa exigencia de control con la que se compromete.

Otro de los rasgos que hay que destacar en la teoría de Dennett es que la responsabilidad es una cuestión gradual, que depende del nivel de razonamiento evaluativo, un meta-nivel que debería propiciar preguntas que revisasen y sometiesen a crítica nuestros objetivos, nuestras pretensiones, nuestros modos de ser y de aquello que queremos llegar a ser. En este punto, Dennett está profundamente influido por Charles Taylor y su artículo «Responsibility for Self», en el que se propone el concepto de auto-evaluación fuerte como criterio para mostrarnos responsables de nuestro modo de ser; en la auto-interpretación y re-evaluación se llega a un punto en el que no hay disponible un metalenguaje, un léxico valorativo de orden superior que sea el criterio de la evaluación, sino que se llega a un momento en el que se pone en juego el tipo de persona que se quiere llegar a ser. La riqueza del léxico valorativo y de la capacidad de auto-evaluación depende también de los contextos socio-políticos en los que se mueve el individuo, de modo que cabría introducir el carácter gradual de la responsabilidad moral, teniendo en cuenta el contexto socio-político.

El propio Dennett comenta esta idea de la ausencia de un meta-nivel de evaluación, pues agudamente señala que una mayor reflexión, un mayor auto-conocimiento no implica necesariamente un "auto-mejoramiento" (*self-improvement*). En todo caso, lo innegable es el mayor grado de responsabilidad por el modo de ser de cada uno.

Recapitulando, la auto-formación en el mundo determinista parece salvarse mediante los niveles de meta-evaluación crítica de que disponemos los seres con un lenguaje articulado. De esta idea, Dennett parece extraer una concepción del yo como centro de gravedad narrativa, pero antes de abordar esta consecuencia, recuperaremos un tema que quedaba pendiente.

En *Elbow Room*, Dennett comenta que esta versión de la idea de acciones autoformativas podría ser vulnerable a la variante indeterminista del argumento escéptico de G. Strawson (en su caso, representada por la versión de P. Edwards). Si ha sido cuestión de suerte que unos tengamos mayor nivel de auto-reflexión, de capacidad de dominio del lenguaje, entonces no todos tenemos las mismas oportunidades de llegar a ser moralmente responsables. La tesis de Dennett es que evolutivamente estamos capacitados para el auto-control y la deliberación, y que el proceso de adquisición de agencialidad (*agenthood*) es como una maratón, y no como un *sprint*, en la que una pequeña desventaja inicial debida a la suerte (por ejemplo, que la salida en la carrera se ordene por el mes de nacimiento) se puede superar. La intención del autor es mostrar que la dicotomía afortunado-desafortunado no es exhaustiva, sino que la fortuna (en este caso cabría hablar de suerte constitutiva) juega un papel importante pero no crucial en la construcción de la personalidad moral.

De este modo, se trata de una minimización del concepto de suerte o fortuna, de acuerdo a la potenciación de la idea de capacidades que poseemos en tanto que seres humanos, unas capacidades que crean unas expectativas necesarias para la vida social. De hecho, Dennett presenta una relación inversa entre capacidades y fortuna.

En definitiva, de lo que se trata es de reivindicar la capacidad de la evitabilidad, de alejarse de situaciones en las que nos ponemos en peligro, en las que nos reivindicamos como agentes dignos de ser responsables. Por decirlo así, Dennett defiende la capacidad de encontrarse en situaciones donde el papel de la fortuna se minimice, y en las que la facultad de racionalidad práctica actúe de acuerdo con la autoevaluación del individuo.

En años posteriores, Dennett extrajo y desarrolló la idea de yo que se podría derivar de su teoría de la acción moral. Para dar cuenta de su concepción del *self*, el filósofo norteamericano recurre a la idea de "centro de gravedad", concepto de la ciencia física, pero aplicado a la narración. Del mismo modo que el centro de gravedad es una abstracción que tiene sentido y con la que se opera en las ciencias físicas, el yo no deja de ser una abstracción que tiene la virtud de ser el sustrato del que se predican diversas propiedades; de hecho, es el resultado de las diversas atribuciones que se nos imputan y que nosotros mismos nos hacemos.

Esta concepción del yo suscita inmediatamente la respuesta crítica, que Dennett concentra en la figura de Otto (un personaje figurado, inventado, narrado, pero cuyo "centro de gravedad narrativa" parece ser el propio Dennett), que respondería lo siguiente: el centro de gravedad no sería más que una ficción del teórico, sin ninguna realidad ni consistencia, con lo cual parece demasiado débil para soportar el peso de la responsabilidad moral. La respuesta de Dennett es tajante: ésa es justamente su gloria, su virtud, pues son ficciones de las que cualquiera estaría orgulloso de haber creado, y pone como ejemplos personajes de ficción (como Ismael, de *Moby Dick*), para mostrar que al referirnos a Ismael no hablamos de Melville, ni del texto, sino de un personaje ficticio. El problema es cómo entender este ficcionalismo: ficción es lo que se ha hecho, creado, inventado, pero también lo fingido, lo que se distancia de lo sucedido. En este segundo sentido es como cabe entender la duda de Otto, el *alter ego* de Dennett, al que éste se dirige cuando señala que «creo que sé dónde quieres llegar. Si un yo no es una cosa real, ¿qué ocurre con la responsabilidad moral?» (Dennett, 1992: 429). Ésta es la cuestión última, a la que da una respuesta abierta,

pues reconoce que «la tarea de construir un yo que pueda *asumir* responsabilidad es un proyecto social y educativo central» (Dennett, 1992: 429-430).

Parece ser que la salida compatibilista de Dennett requiere en último término un determinado proyecto socio-político para garantizar al máximo la creación de individuos responsables, capaces de auto-evaluación y auto-interpretación, capaces de objetivar las razones que les dan y que ellos mismos proporcionan de acuerdo con ciertos criterios de racionalidad.

Sin embargo, en la exposición de Dennett se produce un curioso giro en cuanto a la ontología del yo: en el esquema evolutivo, «eres aquello que controlas y aquello por lo que te preocupas». Con esta sentencia, parece darse una inversión en el siguiente sentido: si uno es lo que controla, y si el control es el criterio de la responsabilidad moral, parece que hay un solapamiento entre lo que uno es y la responsabilidad moral. No obstante, esto supone que estamos constituidos por cosas que controlamos, mientras que parece bastante plausible que estamos configurados, que estamos y hemos sido conformados por acontecimientos que no podríamos controlar. De hecho, en la propia argumentación de Dennett se produce una paradoja que podría dificultar su teoría del yo: las narraciones que nos configuran las producimos no necesariamente de modo consciente y deliberado, de tal manera que no podríamos ser moralmente responsables de acciones inconscientes, como la de crear nuestro propio yo, al menos de un modo que requiriera una ultimidad fuerte en el sentido de Kane. En todo caso, ésta sería una dificultad menor en tanto que el autor demanda un cierto grado de autoconciencia en el proceso de creación de uno mismo.

Indagando en esa dirección, según Dennett, en el proceso de creación de un yo, el sujeto no parte de ninguna identidad rectora, de un Jefe de la Mente que dirija el proceso (Dennett, 1989: 13-14). Más bien lo que ocurre sería una producción inconsciente de diversas identidades, de diversas personalidades, que se van configurando y excluyendo, hasta que una es elegida o surge vencedora de la contraposición con el resto. Este argumento está vinculado a su tratamiento de los desórdenes de personalidad múltiple (MPD), casos de los que trata de dar razón mediante esta descripción del yo como centro de gravedad narrativa. En estos casos, podemos considerar la existencia hay diversos yoes en conflicto, sin que se haya producido un predominio de uno de ellos para darle la unidad mínima requerida al agente para que pueda ser considerado como una persona moralmente responsable.

Una vez presentado el esbozo de la propuesta de Daniel Dennett, todavía nos queda intentar mostrar algunas dificultades que parece presentar la argumentación de nuestro autor.

La primera sospecha escéptica o determinista radical podría ser la siguiente: el lenguaje ha sido resultado de un conjunto de series causales determinadas en la evolución, y más que un instrumento que podamos utilizar para construirnos a nosotros mismos, es una maquinaria automática que nos construye a nosotros. El lenguaje es una cadena que vincula a una sociedad entre sí y con su pasado, de modo que es un instrumento que no podemos improvisar o recrear hasta el punto de garantizar nuestra construcción del yo como una verdadera auto-construcción.

La respuesta del compatibilista acudiría en búsqueda de una intuición cotidiana que pondría en cuestión este problema: el lenguaje es una tradición recibida, pero sometida a crítica, revisión, modificación, recreación, es mudable (inter- e intra-

lingüísticamente), y a su vez abre la posibilidad de distanciarse de los motivos, razones y creencias que podrían transmitirse mediante él.

Sin embargo, hay una versión débil de esta crítica que no plantea el qué del lenguaje, sino el cómo y el quién, y que parece más sugerente y problemática para la posición de Dennett: se trata de la crítica de circularidad. Clark ha señalado que en la constitución narrative del yo se da una alternativa fatal: o bien es falsa o bien es circular. O bien presuponemos una entidad previa que es la que narra la biografía y que poco a poco va quedando modificada hasta ser asumida por la propia narración (fenómeno complejo y poco plausible), con lo que habría una personalidad previa que sería la condición del yo del que quisiéramos ser responsables, pero que quedaría fuera de nuestro control. O bien parece que el yo emerge de una lucha por la existencia entre diversos yoes posibles que han ido gestándose en la actividad narrativa sin sujeto previo, como si el lenguaje fuese un instrumento auto-gestionado. Sin embargo, Dennett explicita en algunos textos que se da una "elección" de ese yo que ha de ser el *Head of Mind*: el problema es, pues, desde qué criterios, desde qué personalidad, se da esa elección. La interpretación más plausible dentro de las coordenadas dennettianas es que se da lo segundo, teniendo en cuenta que el lenguaje es un instrumento democrático, que pueden utilizar todos los yoes que van configurándose en nuestra propia narración y en la que participan otros agentes de manera activa.

Pero además de este problema inicial en cuanto al cómo se produce esa narración del yo, tenemos el problema del resultado, del, por decirlo así, nivel de densidad ontológica del yo. Es decir, ¿está el yo de Dennett tejido de una manera suficientemente fuerte como para resistir el peso de la responsabilidad moral? Esta era la pregunta que Otto, ese oponente ficticio creado por el propio Dennett, le planteaba y a la que el autor respondía que justamente esa aparente debilidad era su gloria y virtud. Sin embargo, esta respuesta no parece demasiado convincente pues depende de toda una ontología de los personajes de ficción que puede estar abierta a discusión. No obstante, hay un indicio que nos puede hacer pensar en la insuficiencia de esa idea del yo: Galen Strawson, crítico y escéptico en cuanto a la idea de auto-configuración del yo (y, por consiguiente, de la responsabilidad moral), acepta y asume la concepción del yo que presenta Dennett. Hay acuerdo entre Dennett y Strawson en cuanto a que no puede haber una verdadera auto-creación, pero está claro que el primero no acepta el tipo de argumento, y ni mucho menos la conclusión que presenta el segundo. Pero si esto es así, ¿cómo es posible que Strawson acepte el concepto de yo de Dennett? Quizá porque a su juicio le sirva como elemento crítico en su estrategia para mostrar la imposibilidad de la responsabilidad moral. Esto nos tendría que indicar que habría que reforzar o dar mayor consistencia al yo narrativo de Dennett, sin convertirlo por ello en un objeto material, es decir, sin transgredir los límites de un materialismo crítico, para poder salvar la idea del compatibilismo y de auto-configuración en un mundo determinista. En esta dirección de reforzar el yo narrativo apuntaría Dennett al hablar del proyecto educativo y social de la creación de individuos, lo cual abre el problema de la cuestión social: «la tarea de construir un yo que pueda *asumir* responsabilidad es un proyecto social y educativo central» (Dennett, 1992: 429-430).

Derivada de este punto aparece la siguiente dificultad: Dennett minimizaba el papel de la suerte y ensalzaba la idea de capacidad en cuanto a la cuestión del auto-creación de los yoes. Ahora bien, ¿hasta qué punto es compatible el "adelgazamiento"

narrative del yo, ese ligero tejido, con la importancia que le da Dennett a las capacidades? Esta es una cuestión que vincula el nivel ontológico del yo con el tipo de cualidades y capacidades que se le pueden atribuir. Un argumento semejante presenta también A. Clark cuando comenta que el tipo de sujeto completo y denso que requiere Dennett para garantizar su capacidad de auto-evaluación radical y de auto-creación del yo (*self-made selves*) no puede garantizar su ontología narrativa porque trata de recurrir a cuestiones culturales, al menos dado el autor al que recurre para inspirarse. El punto que quiere criticar este autor es la cercanía entre Dennett y Charles Taylor, de quien el primero toma la idea del yo como narración en tanto que permite una auto-evaluación; sin embargo, esto supone ya un espacio moral en el que se mueve el sujeto, es decir, es solidario de una cosmovisión comunitarista en la que justamente se da una cierta coordinación entre aquello que uno puede llegar a ser, sus capacidades, sus posibilidades de narración auto-creativa y su posición respecto al bien; es decir, el mapa moral respecto al cual sus acciones van a ser juzgadas. Taylor lo dice en cierto modo cuando señala que «nadie puede ser un yo por sí mismo» (Taylor, 1989: 36), pues estamos en redes de interlocución, y si continuamos con la metáfora de las redes, Dennett insistiría en la procedencia biológico-evolutiva de esas redes y en nuestra capacidad de tejer, destejer y suturar nuestra identidad a partir de ellas. Este supuesto de una cosmovisión comunitarista en el caso de Taylor (acercándose a MacIntyre) es claro en la medida que él mismo polemiza con la tradición que identifica con Locke-Hume-Parfit, que tratarían de buscar lo que él llama un "yo puntual" en el que «el yo es definido en términos neutrales, fuera de ningún marco esencial de cuestiones» (Taylor, 1989: 49). En cierto modo, pues, Dennett comparte con Taylor algunos aspectos como la importancia de la narración de la historia que nos ha llevado a ser como somos, la idea de estar inmersos en una red de interlocutores y la auto-evaluación. Sin embargo, hay una diferencia crucial entre ambos: la idea de yo en Dennett era un *abstractum*, mientras que para Taylor es una sustancia fuerte conformada y orientada por una idea comunitaria de bien que se encuentra en el marco moral como en un mapa. Esta caracterización sustancial de la ontología del yo implica además un peligro para la posición de Taylor, pues la referencia a marcos en los que se forma un sujeto, podría desviar la atribución de responsabilidad y disculpar acciones de los sujetos que se encuentran fuera de los marcos en los que fueron configurados para actuar: de hecho, quizá ni lo fueran en sus propios marcos, porque hay una retro-atribución en dirección al marco (el responsable último es la cultura-marco y no el sujeto). Esto es, hay una tensión inherente en su propia teoría entre la influencia de los marcos y la capacidad de auto-evaluación; si ésta no supera la idea del marco del bien comunitario, entonces la propuesta de Taylor se encuentra con la dificultad de la atribución de responsabilidad a sujetos fuera de su marco cultural de auto-evaluación: es el límite de su cosmovisión comunitarista y multicultural la que corre el peligro, ya en el plano político, de hipostasiar esos marcos culturales que configuran la idea de bien y convertirlos en referentes no sólo de las creaciones culturales sino también de la responsabilidad moral y política. Pero por otra parte, en la teoría de Dennett se aceptaba que el yo era un centro de gravedad, más bien cercano a la tradición Locke-Hume (en quien Dennett reconoce que se inspira). Aquí está en juego la siguiente cuestión: ¿un yo puntual, abstracto, puede llevar a cabo un tipo de auto-evaluación pensada para un yo ontológicamente más denso y referido a un marco cultural? Si

puede llevar a cabo esta auto-evaluación independientemente del marco moral concreto, ¿no será acaso más *profunda* que aquella que se encuentre limitada por el marco de una visión comunitaria del bien?

Hasta cierto punto, aquí resurge la cuestión que Otto le planteaba a Dennett: la ficcionalización o narrativización del yo parecía debilitar, desde el punto de vista del sentido común, el soporte al que referir las atribuciones de responsabilidad moral. Pero por otra parte, y como parece que hemos intuido a partir del comentario de Taylor, un yo puntual, narrativizado o ficcionalizado parecería permitir una auto-evaluación radical mayor que un yo sustantivo limitado por un marco de bien común. Quizá entonces el problema de la ontología del yo y la posibilidad de auto-configurarse en un mundo determinista tenga que moverse entre la Escila de un yo no tan ficcionalizado como para soportar las imputaciones de responsabilidad moral y la Caribdis de un yo no tan "enmarcado" en un proyecto comunitario como para poder llevar a cabo una "auto-evaluación" radical. La propuesta de Dennett (compatibilismo naturalizado) parece no caer en este segundo peligro, más inminente quizá para la posición compatibilista "cultural" de Taylor, pero se encontraría con la dificultad primera, puesto que su idea del yo como centro de gravedad narrativa es aceptada por un escéptico en cuanto a la responsabilidad moral como es Galen Strawson: podríamos decir que aquí Otto, el *alter ego* de Dennett, y Galen Strawson coinciden en que un yo como centro de gravedad narrativa no garantiza el nivel de control y ultimidad que requeriría la noción de responsabilidad moral.

Quizá una perspectiva compatibilista tenga que conjugar el marco naturalista-evolutivo con el socio-cultural de un modo equilibrado para evitar estas dificultades que hemos esbozado. En este sentido, Fischer y Ravizza tratan de establecer las etapas, la historia causal, a través de la cual se produce el entrenamiento moral, la capacitación para llegar a ser considerados seres moralmente responsables. Esta referencia al marco cultural de educación trata de incluir la intuición del marco cultural de Taylor, pero sin llevarlo al extremo comunitarista de este autor. Se trataría de crear un "mecanismo *reflexivo*", aunque en el desarrollo educativo habría que incluir mecanismos no-reflexivos, siempre en orden a conseguir un agente moralmente responsable: la vida en sociedad nos impele a ello. Esta historia de la asunción de responsabilidad abre las puertas a una versión del compatibilismo que se acerca a cuestiones de filosofía política: de hecho, tanto Hurley como Stephen White han apuntado hacia la idea del equilibrio reflexivo de Rawls como punto a tener en cuenta en una versión más amplia de su compatibilismo. En última instancia toda esta cuestión podría remitir a otro pensador de la psicología social que ha sido muy utilizado en filosofía política: Kohlberg. Las etapas del desarrollo moral podrían inspirar una salida compatibilista que tratara de configurar esa historia de los agentes mediante la cual se da una adquisición y entrenamiento de la capacidad de la responsabilidad moral, con la ventaja respecto a Taylor de integrar la cuestión social sin hipostasiar el marco de referencia cultural. Sin embargo, estas ideas requerirían un desarrollo y un balance crítico que exceden el alcance de este texto.

Parece que el complemento que requiere la configuración de un agente capaz de originar sus acciones y decisiones a partir de sus propios deseos y creencias es que dicha originación, dicho principio de autonomía, descanse sobre un lecho de posibles alternativas de acción todavía no realizadas pero que podrían serlo en el momento de

la acción (o elección) del agente: el denominado Principio de Posibilidades Alternativas.

El Principio de Posibilidades Alternativas puede definirse del siguiente modo:

> (PPA) Un agente es responsable si y sólo si dada una situación (y en toda situación *caeteris paribus*) dicho agente pudiera haber actuado de modo diferente a como en realidad actuó.

Para que esta acción pueda ser catalogada como libre y voluntaria el agente, parece, debe ser capaz de poder explicarla a partir de razones que lo movieron a actuar del modo en que lo hizo, razones que actuarían como guías de la acción humana, incluidas dentro del mundo, con poder causal.

Si atendemos a su desarrollo ontogenético, como aquí hemos hecho, la cuestión principal acerca de la responsabilidad moral de un agente es optimista. Porque a pesar de que los deseos y creencias que nos mueven a actuar, según Dennett, son producto de una acumulación e interiorización de normas y convenciones sociales, esto no nos debe hacer desesperar en nuestro intento de actuar libremente dentro de un mundo concebido como determinista (Dennett, 1988).

Pero, ¿cómo es posible que en un mundo dominado por la explicación física de los hechos tenga cabida la acción llevada a cabo desde la vida mental de las máquinas (humanas)? Esto es, si concebimos que todo hecho del mundo tiene una explicación física, ¿cómo podemos atribuir poder causal a lo mental? Esta acusación de epifenomenismo puede serle imputada a Dennett porque parece dotar a la vida mental interna de los individuos de un poder causal que el determinismo deniega. Pero es una imputación falsa.

Dennett concibe al ser humano como intencional; esto es, explica las actitudes comportamentales de un agente desde sus propios propósitos y fines (Dennett, 1971). Aunque dicha intencionalidad es concebida como derivada (Dennett, 1987): lo que los agentes suponen como sus intereses en realidad son fines debidos a la acumulación aleatoria y arbitraria de modificaciones ambientales indicadas por algoritmos que permitieron la configuración de un cierto tipo de orden, una rutina, capaz de crear, desarrollar y evolucionar cierto tipo de organismos que consiguieron elevarse hasta la posibilidad de controlar su propio desarrollo (vista desde el punto de vista filogenético). Es decir, todo agente es un artefacto, incluidos los seres humanos. La única diferencia existente entre la inteligencia humana y algún otro tipo de inteligencia (artificial) es que esta última ha sido creada por la primera, mientras que la inteligencia humana es producto de nuestros genes egoístas (Dennett, 1995). Pero todos responden al mismo principio: su intencionalidad responde a los intereses de quien los diseñó.

Dennett afirma que la cultura humana es producto de la evolución memética. En la evolución del ser humano como organismo biológico, apareció en un determinado momento una serie de mecanismos capaces de adaptarse al medio según los elementos externos (culturales) presentes, mecanismos necesarios para la supervivencia de la especie. Dichos mecanismos fueron incorporados a la estructura fenotípica de la especie, elementos que permiten la acción. Así, lo que en un principio parecían elecciones libres de los individuos, en realidad son respuestas de los mecanismos implemen-

tadores de la acción, mecanismos que a pesar de ser flexibles y adaptables a los contextos, tienen una explicación física. Podemos anular así la acusación de epifenomenismo dirigida a Dennett porque aunque parece que pretende explicar hechos físicos desde lo mental, en realidad tal vida mental no es más que una ilusión edificada por la estructura orgánica (explicable en términos biológicos) del ser humano, que responde a las leyes evolutivas.

La dificultad añadida surge cuando se pretende la negación del mencionado PPA. Tendemos a pensar que somos libres y que en un momento dado podríamos haber elegido actuar de modo diferente a como realmente actuamos, por lo que cuando no podemos hacer otra cosa consideramos que se ha dado un cierto fatalismo local (un cierto tipo de determinismo puntual) que impide la posibilidad de actuar de otro modo, una especie de coacción que imposibilita cualquier alternativa.

Dennett niega el PPA porque 1. no existen evidencias que indiquen que una persona podría, o puede, haber hecho otra cosa que lo que realmente hace y 2. porque la responsabilidad del agente respecto de la acción no remite aunque éste no pudiera haber actuado de otro modo. Pero Dennett va todavía más lejos: no podemos actuar de otro modo ni siquiera en condiciones similares. Las clausulas *caeteris paribus* no funcionan porque en la mayoría de los contraejemplos las condiciones de acción se ven alteradas enormemente, porque las condiciones cambian en cada circunstancia. Nunca podemos estar dos veces en la misma circunstancia porque nuestro diseño como agentes (nuestros contenidos) cambia. El individuo aprende, se aburre, deja de lado… y cada situación supone un cambio en los contenidos que lo mueven a actuar. No es posible poder actuar de otro modo en las mismas circunstancias porque las mismas circunstancias nunca se repiten y nunca nadie ha podido hacer otra cosa que aquello que realmente hizo. El contenido mental de los hombres se modifica con cada suceso. Ya seamos deterministas o seamos indeterministas, nuestra estructura cambia constantemente. Nunca estamos en la misma situación. Que uno pudiera o no actuar de otro modo en la situación dada es irrelevante: no pudimos hacer otra cosa que lo que realmente hicimos.

Parece que la diferencia entre poder haber hecho otra cosa y no poder haberla hecho es el determinismo, pero no es así, la responsabilidad es exactamente la misma si estamos determinados como si no lo estamos, porque en un momento dado nosotros hicimos lo que hicimos por nuestro diseño. Actuamos como actuamos porque tenemos los contenidos mentales que tenemos, pero ello no supone una reducción de la responsabilidad de nuestra acción porque responde a nuestra propia historia cognitiva, la acción nos pertenece por estar causada por los contenidos cognitivos del agente. Es una ilusión pensar que si estamos determinados no somos responsables de lo que hacemos porque es nuestra estructura la que causa las acciones que realizamos y nuestra estructura depende de nuestra historia cognitiva y del modo en que consideremos las enseñanzas. Para Dennett, el PPA es una ilusión porque no podemos nunca hacer otra cosa que la que realmente hicimos, pero no por ello dejamos de ser responsables de lo que hacemos. Parece que la solución para la toma de decisiones es tener mayor control, ser más racional, contener más normas, más justificaciones… en pocas palabras: debemos estar diseñados para poder actuar del modo en que debemos.

Los argumentos de Dennett parecen entroncar con otro argumento similar: el argumento de la consecuencia de Peter van Inwagen. Este autor, que explícitamente

niega la posibilidad de la coexistencia del determinismo y de la libertad de acción, considera que es posible la libertad del ser humano, independientemente de que existamos en un mundo determinista o indeterminista. Lo que nos viene a decir van Inwagen es que si el mundo es determinista, el agente no tiene ninguna opción diferente de actuar más que aquella que realizó, y que toda vez que se repitiera (pongamos por caso que un genio hiciera retroceder el tiempo indefinidamente hasta el instante anterior a la acción) el agente actuaría exactamente del mismo modo, porque sobre él actúan unas leyes físico-causales que imposibilitan cualquier alternativa. Pero el agente no lo tendría mejor en un mundo indeterminista: en tal caso, el agente se vería obligado a actuar de una determinada forma todas las veces que la situación se repitiera, aunque podría haberse dado el caso de verse obligado por algún otro curso de acción que ya no está disponible, porque el agente está sometido a una serie de leyes indeterministas que deja abiertas varias vías de acción pero que, una vez actualizada una de ellas, clausura las demás.

Lo que tienen en común los argumentos de Dennett y van Inwagen es que ambos suponen la existencia de una causa completa para toda acción. Es decir, suponen que dada una determinada acción existe un, y sólo un, modo factible de explicación. Pero, entonces (y es lo importante para nuestro tema), si todo hecho físico, en este caso una acción humana, puede ser explicado a partir de una causa completa (que debe ser dada en términos físicos), entonces cualquier pretensión de causalidad de lo mental queda anulada.

Cuando negamos la posibilidad de actuar de otro modo distinto a como en realidad actuamos, damos por supuesto que no tenemos otra salida que realizar la acción que hacemos; esto es, que estamos determinados. Cuando hablamos de acciones correctas no importa en absoluto que no pudiéramos hacer otra cosa, porque damos por supuesto que lo que cabe hacer en dicha situación es lo que haría cualquiera de nosotros en el mismo (o similar) caso. Pero no ocurre así cuando evaluamos acciones que consideramos incorrectas.

Cuando por lo que nos preguntamos es por la censura de un cierto acto, si que tiene importancia el poder llegar a determinar que un cierto agente no tenía más opciones que realizar el acto que definitivamente hizo. Porque si llegáramos a la conclusión de que el agente tenía otras posibilidades de acción, entonces cabe la posibilidad de que el agente se equivocara al hacer lo que hizo y, por lo tanto, ser un acto censurable. Si, por el contrario, demostramos que no es posible que el agente actuara de modo diferente, entonces demostramos que el agente estaba obligado a actuar como lo hizo y desaparece (o, por lo menos, disminuye) la responsabilidad de sus actos.

Agradecimientos

Este trabajo se ha realizado parcialmente dentro del proyecto “Creencia, motivación y verdad” (BFF2003-08335-C03-01).

Referencias

1. Dennett, D.C.: Brainstorms: philosophical essays on mind and psychology, Montgomery (1978)
2. Dennett, D.C.: Elbow Room. The Varieties of Free Will Worth Wanting, Oxford U. Press, Oxford (1984)
3. Dennett, D.C.: I could not have done otherwise –so what?, The Journal of Philosophy, 81, (1984) 553-567
4. Dennett, D.C.: The intentional stance, MIT Press, (1987)
5. Dennett, D.C.: The Origins of Selves, Cogito, 3, (1989) 163-173
6. Dennett, D.C.: The Self as a Center of Narrative Gravity, in Kessel, F.; Cole, P. and Johnson, D. (eds.): Self and Counsciousness: Multilple Perspectives, Erlbaum, Hillsdale (1992)
7. Dennett, D.C.: Consciousness Explained, Penguin, London (1992)
8. Dennett, D.C.: Darwin's dangerous idea: evolution and the meaning of life, Simon and Schuster, (1995)
9. Dennett, D.C.: How to do other things with words, Royal Institute Conference on Philosophy and Languague, in Preston, J. (ed.): Philosophy, 42, supplement (1997) 219-235
10. Dennett, D.C.: Freedom evolves, Viking Penguin (2003)
11. Dennett, D.C.: On failures of freedom and the fear of science, Daedalus (2003) 126-130
12. Dennett, D.C.: Making ourselves at home in our machines, Journal of Mathematical Psychology, 47 (2003) 101-104
13. Dennett, D.C.: Natural Freedom (Reply to A. Mele, J.M. Fischer and T. O'Connor), Metaphilosophy, 36/4 (2005) 449-459
14. Dennett, D.C. and Hofstadter, D.R. (eds.): The Mind's I. Fantasies and Reflections on Self and Soul, Penguin, London (1981)
15. Strawson, G.: Freedom and Belief, Oxford U. Press, Oxford (1986)
16. Strawson, G.: The Self, Journal of Consciousness Studies, 4/5-6 (1997) 405-428
17. Van Inwagen, P.: An essay on Free Will, Clarendon Press, Oxford (1983)

Formalización del lenguaje filosófico en Leibniz[1]

Leticia Cabañas

IES Victoria Kent – Torrejón de Ardoz (Madrid)
lcabanas@telefonica.net

Como muchos de sus contemporáneos buscó Leibniz una filosofía renovada –*philosophia reformata*– más allá de la semibárbara escolástica y de la *secta machinalis* de los Modernos. Desde el principio mostró interés en hacer de la filosofía un conocimiento seguro para, mediante la mejora de los métodos, alcanzar "une philosophie demonstrative"[2], según su plan genial de abarcar demostrativamente todo el saber. Su proyecto filosófico apunta a hacer explícitas las condiciones de inteligibilidad a la base de la complejidad estructural de un universo armónico e intrínsecamente racional, en donde todos los hechos están interconectados. La construcción de un lenguaje científico o *lingua philosophica* permitirá una correcta interpretación de la realidad fundada sobre bases objetivas. Una lengua racional que sirva al perfeccionamiento de las funciones de la mente, un *organon* de la razón con el que superar nuestras limitaciones cognitivas naturales, detectando los errores de modo automático e infalible hasta el punto de hacerlos inexpresables. Un auténtico telescopio de la mente que extienda el alcance de nuestra facultad cognitiva para descubrir la verdad y eliminar el error.

En este sentido, rechaza Leibniz el conocido argumento de Descartes contra la utilidad de una "lengua filosófica" para el avance del conocimiento. En respuesta a Mersenne del 20 de noviembre de 1629 aducía Descartes que tal lengua no puede hacer avanzar la ciencia, puesto que depende de ella, quedando subordinada la construcción de una lengua artificial a alcanzar la *vera philosophia*, objetivo fundamental de la investigación cartesiana[3]. Para construir una lengua científica primero hay que poseer la totalidad del conocimiento, la *mathesis universalis* –término de herencia

---

1 Siglas utilizadas:

A Leibniz, G.W.: *Sämtliche Schriften und Briefe*. Hg. von der Akademie der Wissenschaften (Akademieausgabe). Reihe I-VII. Darmstadt, später Leipzig, zuletzt Berlin 1923 ff.

GP Leibniz, G.W.: *Die philosophischen Schriften von G.W. Leibniz*. Hg. Carl Immanuel Gerhardt. 7 Bände. Berlin 1875-1890 (Reimpresión: Hildesheim-New York 1978).

GM Leibniz, G.W.: *Leibnizens mathemathische Schriften*. Hg. Carl Immanuel Gerhardt. 7 Bände. Berlin (später Halle) 1849-1863 (Reimpresión: Hildesheim-New York 1971).

C Leibniz, G.W.: *Opuscules et fragments inédits de Leibniz*. Extraits des manuscrits de la Bibliothèque royale de Hanovre par Louis Couturat, Paris 1903 (Reproducción : Hildesheim – New York 1971).

Bodemann *Die Leibniz-Handschriften der Königlichen öffentlichen Bibliothek zu Hannover*. Hannover und Leipzig 1889. (Reimpresión: Hildesheim 1966).

NE *Nouveaux Essais.*

GI *Generales Inquisitiones.*

2 GP IV, 347.

3 Descartes, *Oeuvres*, ed. Adam et Tannery, París, 1982-1991, vol. I, p. 81.

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 174-185, 2006.

platónica que emplea Descartes en la 4ª Regla– por la que a partir de la lista completa de las ideas elementales nombradas y simbolizadas, todo lo pensable es reconstruible mediante un cálculo. No hay que decir que el filósofo francés se muestra escéptico sobre la posibilidad de llevar a término tal programa, considerando utópica la empresa. Por su parte Leibniz tiene una visión más pragmática respecto a la auténtica dimensión del lenguaje racional. En un breve comentario a la carta de Descartes a Mersenne argumenta que para realizar la lengua universal no es necesario haber alcanzado el grado sumo de conocimiento de la verdad, pues aunque dicho lenguaje dependa de la auténtica filosofía, no depende sin embargo de su perfección. Podemos construirlo aunque la filosofía no logre su completud y en la medida en que la ciencia se desarrolle también lo hará el lenguaje[4].

Los creadores de lenguajes universales que preceden a Leibniz siguen una línea que va de Lulio a Kircher y a los ingleses Dalgarno y Wilkins, partiendo estos últimos de las indicaciones ofrecidas por Francis Bacon en el *Advancement of Learning* y en la *Instauratio Magna*. Nuestro autor se muestra crítico frente a todas las propuestas anteriores de lenguaje universal, pues opina que el sistema de caracteres que proponían es arbitrario, con ausencia de un fundamento semántico extrínseco, que sus sistemas simbólicos no se construyen sobre un análisis de los pensamientos[5]. Sabe Leibniz que la renovación de la especulación filosófica depende de una reorganización del lenguaje científico. De acuerdo con el ideal aristotélico de un perfecto saber e implicado en la lucha durante el Racionalismo por lograr un conocimiento más seguro, tiene como preocupación primera la de construir un lenguaje completamente formal, en línea con la meta pansófica de un lenguaje que exprese todo posible conocimiento humano, la omnisciencia prometida por los Rosacruces. Un lenguaje artificial o *Characteristica Universalis* que sobre la base de una teoría de signos opere formalmente exponiendo los procedimientos lógicos que articulan las proposiciones y sus relaciones. Esto frente a pensadores contemporáneos, como Locke, que no tiene en absoluto intención de preocuparse por el orden formal, pues cree que es imposible reducir a reglas los diversos grados en que los hombres asienten sobre algo. Y como anteriormente Descartes que –al igual que Bacon– considera estéril el lado formal de la lógica, una pesada cadena de la que hay que deshacerse si se quiere verdaderamente contribuir al progreso de la ciencia. No ve valor en el desarrollo de un cálculo lógico, lo que provoca la frase lapidaria de Leibniz: "Des Cartes... avoit l'esprit assez borné"[6], pues al no tratar la lógica formal limita el saber humano.

En un siglo XVII donde las reglas silogísticas estaban pasadas de moda pero se mantenía la creencia en que la lógica de Aristóteles constituía la forma definitiva, Leibniz se interesa por aclarar las relaciones que están a la base del sistema lógico y extender la noción de forma más allá del silogismo. Para construir la lógica sobre nuevas bases se desmarca de la lógica tradicional, un método rudimentario que es

---

[4] "Cependant quoyque cette langue depende de la vraye philosophie, elle ne depend pas de sa perfection. C'est à dire cette langue peut estre établie, quoyque la philosophie ne soit parfaite: et à mesure que la science des hommes croistra, cette langue croistra aussi", C 28.

[5] *A Burnett*, 24 agosto 1697, GP III, 216.

[6] GP IV, 297.

posible superar[7], proponiendo una importante modificación del modelo demostrativo axiomático. No es Leibniz, sin embargo, el único pensador que supera el mediocre nivel general de la lógica del siglo XVII: la *Logica Hamburgensis* de Jungius y la *Logique de Port-Royal* de Arnauld y Nicole son también tentativas por enriquecer la lógica con nuevas formas de razonamiento irreductibles a la silogística. Pues entienden que el marco lógico de la ciencia clasificatoria de la Antigüedad debe dar paso a los procedimientos analíticos y matemáticos que caracterizan al pensamiento científico moderno. Pero Leibniz es consciente de que se adelanta a su tiempo, que está solo frente a los esfuerzos por desarrollar una lógica formal, y destina su primer trabajo plenamente original de juventud, el *De arte combinatoria* a completar la lógica de Jungius y la de los lógicos de Port-Royal que, frente a él, temen la desnaturalización producida por el simbolismo y construyen una lógica a partir de un lenguaje natural. Es el primer ensayo de lógica simbólica, un texto complejo por su estilo y por su contenido rico en ideas embrionarias que desarrollará a lo largo de su carrera filosófica y donde define por primera vez el régimen metodológico que gobierna la construcción del lenguaje universal, con una sintaxis que refleje las relaciones lógicas de los conceptos.

La combinatoria leibniziana se diferencia también fundamentalmente del método de Descartes, el cual queriendo romper con el formalismo escolástico adopta posiciones intuicionistas tomadas de la tradición agustiniana que descansan sobre la experiencia del *Cogito*, esforzándose por reducir la deducción a una intuición continuada. Leibniz interpreta esta teoría de la verdad basada en la evidencia, en la claridad y distinción de las ideas –*clare et distincte percipere*–, como un principio epistemológico subjetivo y, por lo tanto, insuficiente. Busca reemplazar el análisis cartesiano por otro que sustituya todo elemento intuitivo por elementos lógicos que no impliquen más que el entendimiento, dejar de lado todo recurso a la intuición para, a continuación, obtener demostraciones exclusivamente combinatorias. Se trata de un formalismo que permite reemplazar operaciones puramente conceptuales por manipulaciones empíricas de símbolos o caracteres que expresan las relaciones entre los objetos, y que son manejados más fácilmente que aquello que representan[8]. Aunque el signo mismo no es empírico: es cierto que se le percibe, como *notam visibilem cogitationes repraesentantem*[9]*,* pero no está sometido a lo confuso de la percepción, pues no es su materia lo que cuenta, sino que su valor reside en las reglas de conexión que muestra. Rescata así Leibniz la dimensión objetiva del signo lingüístico, el aspecto icónico del sistema discursivo, empleando el término *characteres* para designar signos o símbolos de cualquier género: voces, palabras escritas, figuras geométricas, cifras, imágenes[10]. Frente a Descartes, propone acudir a los símbolos lo más posible, por ofrecer a

[7] *A Wagner*, GP VII, 520. Cfr.: "*Posterius obstaculum est imperfectio Artis Logicae*. ita enim sentio, Logicam quae habetur in Scholis, tantum abesse à Logica illa utili in dirigenda mente circa veritatum inquisitionem, quantum differt Arithmetica puerilis ab Álgebra praestantis Mathematici", C 419.

[8] "*Characteres* sunt res quaedam, quibus aliarum rerum inter se relationes exprimuntur, et quarum facilior est quam illarum tractatio", *Characteristica Geométrica*, 10 agosto 1679, GM V, 141.

[9] Bodemann, 80.

[10] "...sive illi characteres sint verba sive notae, sive denique imagines", GP VII, 31.

la mente un soporte sensible, especie de hilo mecánico *–filum meditandi–* que ayuda a la imaginación a sobrepasar sus límites. Al no estar nuestra atención ni nuestra memoria en situación de seguir todas las particularidades y de recordar con orden las varias secuencias del razonamiento[11], es la visualización simbólica del pensamiento la que amplía la posibilidad de su control, concentrando una cadena compleja de pruebas en una única fórmula. Si queremos evitar errores y comprobar la corrección de un razonamiento es necesario recurrir al uso de los caracteres o signos, que brindan la ventaja cognitiva de aliviar la memoria, al no ser necesario recordar datos particulares sino únicamente seguir una regla de conexión.

El Barroco representa el triunfo del autómata como modelo epistemológico. Persigue Leibniz también el automatismo por medio de la manipulación de símbolos, en donde el *filum* que es la forma lógica nos conduzca mecánicamente sin necesidad de esfuerzo mental. El lenguaje formalizado, la C*haracteristica Universalis*, construcción de estructuras de pensamiento que descansan en un sistema de signos, cuenta con la *cogitatio caeca sive symbolica*, el mecanismo como *cogitatio compendium* que permite mantener nuestra mente libre del peso conceptual. Poseen por tanto los signos un papel no subsidiario, sino constitutivo del pensamiento: "si characteres abessent, nunquam quicquam distincte cogitaremus, neque ratiocinaremur"[12]. Sólo tendremos una idea de algo cuando logremos representar el objeto mediante signos, en cuanto que las operaciones con los caracteres son isomórficas a las relaciones lógicas entre los conceptos correspondientes. El *ars characteristica* se presenta como el arte de construir signos y de ordenarlos de modo que mantengan entre sí las mismas relaciones que los pensamientos que representan. Tal analogía constituye el núcleo de todo el sistema formalizado de signos y es gracias a la lógica como conoceremos la estructura racional del mundo, en un paralelismo coincidente con la estructura de las actividades mentales.

Influido por el desarrollo prodigioso de las matemáticas en el siglo XVII, Leibniz tiene como objetivo primero llegar a la computabilidad del lenguaje lógico, aritmetizar todos los contenidos del pensamiento, pues, como anteriormente Descartes, ve en las matemáticas el paradigma de una correcta y eficiente argumentación para la construcción de un sistema coherente. Sin embargo ambos toman las matemáticas como modelo de forma muy diferente, pues mientras que Descartes funda la certeza matemática en la intuición Leibniz lo hará en la evidencia simbólico-formal. En su deseo de trasladar el método matemático a la lógica –conexión estructural que nadie en su época vio tan claramente como él– alaba a Aristóteles por ser el primero en haber argumentado matemáticamente fuera de la matemática[13]. Y a pesar de la crítica fundamental que dirige al materialismo de Hobbes, hace suyas muchas de sus teorías, como la de que *razonar es calcular*, donde la argumentación racional equivale a una operación matemática. Pero frente a la afirmación hobbesiana de que "ratiocinari...

[11] *De mente, de universo, de Deo*, diciembre 1675, A VI, 3, 462.

[12] *Dialogus*, agosto 1677, A VI, 4 A, 23.

[13] "...Aristoteles...der erste in der that gewesen, der mathematisch auβer der *Mathematik* geschrieben",
A *Wagner*, GP VII, 519.

idem est quod addere et substrahere”[14], adelanta Leibniz que el análisis del pensamiento es mucho más complejo que su mera traducción en una adición y sustracción.

El traslado del método matemático a la lógica se efectúa en la citada *Dissertatio de arte combinatoria* (1666), una de las primeras y más fecundas propuestas de cálculo combinatorio en la historia del pensamiento matemático, donde se desarrolla la idea de una *Characteristica Universalis* o doctrina general de los signos, *verum organon Scientiae Generalis* que explota al máximo la mencionada visión hobbesiana de que todo nuestro razonamiento es en última instancia una forma de cálculo. Leibniz atribuye una gran importancia a la representación simbólica de las interrelaciones de identidad y de inclusión en nuestro pensamiento que hacen visible la estructura lógica de las proposiciones en el proceso demostrativo y extiende la *Characteristica Universalis* a todo el ámbito del saber[15]. En la búsqueda del fundamento del orden predicamental de los conceptos, a partir de un alfabeto de signos lingüísticos o caracteres simples que representan conceptos primeros irreductibles, se busca reducir todos los conocimientos humanos a un pequeño número de principios o categorías de aplicabilidad universal capaces de expresar, mediante combinaciones de figuras, todas las relaciones posibles entre conceptos[16]. El método combinatorio o *Ars Combinatoria* se ajusta a un orden de complejidad creciente que no sigue un proceso deductivo lineal, como el cartesiano, sino multilineal, en red, conformando una pluralidad de caminos deductibles posibles que permiten llegar a un punto dado por más de una vía.

En el importante escrito *Generales Inquisitiones* de 1686, el mismo año en que escribe Leibniz el *Discours de Métaphysique,* trata el tema de los fundamentos de la forma lógica. Buscó desarrollar la tesis de Descartes de que todo nuestro pensamiento descansa últimamente en unos conceptos fundamentales innatos, transportándola a su sistema simbólico basado en ideas primitivas, las *notiones absolute primae,* el atomismo conceptual en que se apoya su lógica. Razonar *more geométrico* implica la formalización del lenguaje mediante rigurosas cadenas de definiciones y la referencia a axiomas básicos. La realidad, el mundo de las mónadas, posee una estructura inteligible que puede ser expresada en términos de combinación de conceptos simples o atómicos. Se trata entonces de determinar mediante una combinatoria metódicamente regulada todas las ideas primitivas en que nuestros conceptos pueden resolverse y elaborar un exhaustivo inventario de los términos simples irreductibles[17].

Pero ese plan primero de establecer una lista completa de las ideas primitivas en las que nuestros conceptos se resuelven presenta dificultades insuperables. Duda Leibniz de modo creciente en que esos conceptos primitivos absolutamente simples, elementos últimos no analizables, sean accesibles, que sea posible establecer de forma definitiva el A*lphabetum cogitationum humanarum* integrado por un pequeño número

---

[14] Hobbes, *De corpore*, I, 1, 2.

[15] “L’esprit humain ne sçauroit aller fort avant en raisonnant, sans se servir des caracteres...”, *A Mariotte*, julio 1676, A II, 1, 271.

[16] Este proyecto del *Alphabetum Cogitandi* fue concebido por Leibniz a los 18 años sobre el modelo de Lulio de un *alphabetum artis generalis* integrado por 45 conceptos como elementos de toda ciencia posible.

[17] “Tametsi infinita sint quae concipiuntur, possibile tamen est pauca esse quae per se concipiuntur. Nam, per paucorum combinationem infinita componi possunt”, *De Organo sive Arte Magna cogitandi,* 1679, AVI, 4 A, 158. Cf. GP IV, 72-73.

de pensamientos. Es cierto que el hombre debe luchar con toda su energía por alcanzar los últimos límites del conocimiento, llegar a los conceptos fundamentales inanalizables o *protonoemata*[18], pero tampoco hay que olvidar la irreductible asimetría entre el intelecto divino y el humano. Ya en los años anteriormente inmediatos a su estancia en París comienza a aparecer el problema de la imposibilidad teórica, debido a los límites del entendimiento humano, de obtener términos que sean realmente primitivos e irresolubles. De modo que si bien inicialmente, en su *Dissertatio de arte combinatoria*, asumió el número de conceptos simples como siendo finito, enseguida lo consideró como infinito, reconociendo que su idea original de un alfabeto de los pensamientos humanos había sido demasiado ambiciosa. Y en 1676, en París, habla del más modesto deseo de un catálogo de los pensamientos humanos, pues se da cuenta de la incompletud e impractibilidad de su programa de una *Characteristica Universalis*, no siendo siempre posible encontrar el término de una serie de razonamientos[19].

Acaba Leibniz por admitir no ideas absolutamente primitivas, sino sólo relativas, provisionalmente asumidas[20]. Renuncia por tanto a perseguir los *protonoemata simpliciter* u objetos de pensamiento simples en sí mismos y reconoce que no es necesario suponer tales simples. El fundamento no son ya las *notiones absolute primae*, los conceptos absolutamente primitivos, sino más bien las *notiones quoad nos primae* o conceptos primitivos para nosotros. Habrá que comenzar por un número de términos asumidos como primitivos que para nosotros, en el momento presente, no sean descomponibles y de ellos derivar todas sus posibles consecuencias[21]. Es decir, no se trata de construir un aparato conceptual estático, sino de desarrollar un proceso metodológico que permita el uso óptimo de aquellas nociones con las que contamos dadas las circunstancias. Quedando abierta la posibilidad de que en un momento posterior los conceptos puedan seguir siendo analizados y nos acerquemos así a los conceptos absolutamente simples[22].

---

18 "Conceptus primitivus est, qui in alios resolvi non potest...", *Introductio ad Encyclopaediam arcanam,*1683-85 ?, A VI, 4 A, 528.

19 "Ayant le catalogue des pensées simples on sera en estat de recommencer *a priori*, et d'expliquer l'origine des choses prise de leur source d'un ordre parfait et d'une combinaison ou synthese absolument achevée. Et c'est tout ce que peut faire nostre ame dans l'estat ou elle est presentement", *De la sagesse,* 1676, A VI, 3, 672. Cf. H. Burkhardt, "The leibnizian *Characteristica Universalis* as a link between Grammar and Logic", en: *Speculative Grammar, Universal Grammar, and Philosophical Analysis of Language*, ed. D. Buzzetti / M. Ferriani, John Benjamins, Amsterdam, 1987, p. 46.

20 "*...Terminos integrales primitivos simplices* irresolubiles vel pro irresolubilibus assumtos…", GI, AVI, 4 A, 742.

21 "Je trouva donc qu'il y a des certains termes primitifs si non absolument, au moins à nostre egard…", *Projet et Essais pour avancer l'art d'inventer*, 1688-90 ?, A VI, 4 A, 964.

22 S. Gensini, "Terminus: Scientific Language vs. Ordinary Language from Galileo to Leibniz", en: *Iter Babelicum. Studien zur Historiographie der Linguistik. 1600-1800*, ed. D. Di Cesare / S. Gensini, Nodus Publikationen, Münster, 1990, p. 41; Cf. F. Schupp, en: *Leibniz: Die Grundlagen des logischen Kalküls*, latín-alemán, ed., trad., y notas de F. Schupp, Meiner, Hamburgo, 2000, p. 168.

Finalmente, hacia 1690 abandona definitivamente las investigaciones filosóficas fundadas sobre tentativas de reducción al cálculo o teoría de conceptos. Ya no insiste en la existencia de unas verdades primeras o proposiciones inmediatas, ciertas y autoevidentes, último apoyo de la justificación cognitiva. Se da cuenta de que la teoría analítica conduce a dificultades, contradicciones y discrepancias, que es incapaz de aportar una solución definitiva al problema de la racionalidad. El análisis leibniziano no puede resolverse en un proceso cartesiano de resolución de lo complejo en lo simple, pues tal conocimiento deductivo desde todos complejos hasta los más simples elementos no es la descripción adecuada de un método seguro de conocimiento. Deja Leibniz de ser racionalista en el sentido de creer que el método correcto en metafísica es empezar con axiomas conocidos intuitivamente con la esperanza de derivar de ellos la mayor parte de las verdades mediante una deducción rigurosa. La racionalidad no consiste en la reducción total a un fundamento único que favorezca sólo la vía analítica y el saber de las nociones[23]. Y a la vez que rechaza la metodología racionalista, busca sustituirla por un método de progresar en ciencia liberado de los excesos de la deducción y que eluda la exigencia escéptica de mostrar unos primeros principios indudables. Si en un principio Leibniz consideró que toda su filosofía era deductiva, posteriormente se dio cuenta de que es imposible producir conocimiento sólo con el principio de identidad o contradicción. El entero campo existencial es irreductible a un sistema lógico-formal. Renuncia a proseguir la problemática que preside el argumento de las *Generales Inquisitiones* y en lugar de un análisis que persiga los elementos primitivos como base de una reconstrucción de lo real, ambicionando una síntesis total, prefiere la construcción de herramientas de una eficacia limitada pero útil[24]. Es decir, que desde una orientación racionalista primera, tiende luego a soluciones pragmáticas. Para hacer realidad esa buscada *Scientia Generalis* deberá enfrentarse con la ardua tarea de renovar la forma, el método y la organización del saber de la época, desarrollando una investigación categorial de nuevos fundamentos para la ordenación del conocimiento científico como alternativa al racionalismo modelado en la pura geometría.

Pero a lo que no quiere renunciar es a su doctrina de la verdad como inhesión del predicado en el sujeto, fundamento de la sintaxis lógica, a pesar de la dificultad en admitir el carácter analítico de las proposiciones singulares verdaderas, por ser su resolución una tarea infinita, incluso para Dios. Pues si bien la resolución al infinito del concepto individual es *a priori* para Dios, no puede serlo sin embargo para un espíritu finito[25]. Tal principio escolástico de *inesse* que se remonta a Aristóteles había sido válido desde la Edad Media sólo para los enunciados necesarios. Leibniz lo amplía en su validez de modo que incluya los enunciados contingentes, pues para él la inclusión del predicado en el sujeto no es un aspecto diferencial entre lo necesario y lo contingente: aplica un tratamiento intensional también a las proposiciones singula-

---

[23] S. Brown, "Leibniz's Break with the Cartesian *Rationalism*", en: *Philosophy, Its History and Historiography*, ed. A. J. Holland, Reidel, Dordrecht, 1985, p. 205.

[24] D. Berlioz, « Leibniz ingeniero: el espíritu de invención », en: *Actas del Congreso Internacional Ciencia, Tecnología y Bien Común: La Actualidad de Leibniz,* ed. A. Andreu / J. Echeverría / C. Roldán, Universidad Politécnica de Valencia, Valencia, 2002, p. 236.

[25] Verdades contingentes "...a sola Mente infinita a priori intelligitur, nec ulla resolutione demonstrari potest..." , *De natura veritatis*...1685-86 ?, A VI, 4 B, 1517.

res. La ciencia demostrativa no se limita en Leibniz al ámbito de los universales o de las sustancias segundas aristotélicas, sino que llega también a los individuos o sustancias primeras. La misma definición general de la verdad se aplica a ambos tipos de proposición, al estar basados en una misma forma de relación lógica: la demostración finaliza en la verdad idéntica; la identidad es su último fundamento y a partir de ella ya no hay nada más que probar[26]. Afirma con determinación la analogía en este sentido entre los enunciados contingentes y necesarios, lo que constituye una gran originalidad y genial invención, además de ser uno de los pilares en que se fundamenta el sistema leibniziano. Lo que se intenta es evitar una fractura entre la estructura demostrativa de las disciplinas racionales y de las empíricas que tienen sus raíces en la analiticidad infinita, asegurar la unidad del conocimiento teórico y sensible que mantenga la vieja idea de una sistematización del conjunto de las ciencias. Pues las verdades fácticas no están menos fundadas en la razón que las verdades de razón mismas. Su característica distintiva, la contingencia, es ontológica más que epistemológica.

Leibniz desea una combinatoria total, abarcadora no sólo de una lógica de la necesidad, que resulta insuficiente, sino también de una lógica que dé cuenta de la diversidad del mundo, es decir, que no excluya las proposiciones singulares y que, por tanto, no dependa de un análisis último de los conceptos; en definitiva que siga un proceder diferente al de los sistemas axiomáticos. Se trata de ampliar el análisis finito de las verdades necesarias con el análisis infinito de las verdades contingentes, fácticas. Fueron sus reflexiones en geometría y en análisis infinitesimal las que le permitieron comprender que las nociones son también analizables hasta el infinito, aplicar la metáfora del "análisis infinito" a las proposiciones contingentes permitiendo su racionalización. El descubrimiento y la puesta a punto del cálculo infinitesimal produce un nuevo modelo de identidad. Leibniz justifica la introducción de esta nueva matemática de los infinitesimales, vital para su filosofía, porque incrementa nuestro conocimiento de la realidad. Por influencia neoplatónica considera cada sustancia individual como un microcosmos en el que puede leerse la entera secuencia del universo. El cálculo infinitesimal hace posible que una noción completa de una sustancia cuyos predicados son en número infinito (en cuanto que extiende su acción sobre el resto de sustancias en imitación a la omnipotencia del Creador, comprendiendo todo lo que sucede en el mundo, lo que fue y lo que será, aunque desde la limitada perspectiva de un individuo)[27] sea objeto de un tratamiento formal por un entendimiento finito y desarrollar un discurso inteligible sobre una entidad de naturaleza infinita[28].

El cálculo infinitesimal conecta con la ley de continuidad o "principio del orden general" que gobierna la función clave del infinito en la metafísica de Leibniz y en base al cual la realidad, a pesar de sus diferencias, puede ser vista como un todo ar-

---

[26] "*Demonstrare propositionem* est resolutione terminorum in aequipollentes manifestum facere quod praedicatum aut consequens in antecedente aut subjecto, contineatur", *Praecognita ad Encyclopaediam...,* 1678-79 ?, A VI, 4 A, 135; Cf. "Vera autem propositio est cujus praedicatum continetur in subjecto...", C 401.

[27] *Discours de Métaphysique* § IX, A VI, 4 B, 1541-42.

[28] J.G. Rossi, "Sur deux types de rapport entre sujets et prédicats dans la philosophie leibnizienne", *Studia Leibnitiana.* XXIX/1 (1997), p. 105.

mónico. Al permitir el salto de un orden a otro hace que casos heterogéneos sean entendidos como perteneciendo al mismo principio formal. El principio de continuidad –con el que Leibniz trató siempre, desde su temprana juventud hasta una edad avanzada– aporta orden a la diversidad y los medios para conectar secuencias de determinaciones empíricas con el fin de formar una totalidad teórica. Se pasa ahora del concepto de relación entre elementos discretos al concepto de función entre variables, la asociación por correspondencia biunívoca de elementos que pertenecen a conjuntos diferentes. Si "la difference de deux cas peut estre diminuée au dessous de toute grandeur donnée" [29] se habrá demostrado que la proposición contingente es verdadera a pesar de no ser perfecto el análisis. Aunque nunca pueda mostrarse una coincidencia entre el sujeto y el predicado, jamás surgirá una contradicción. El infinito no podrá ser totalizado, pero la inteligibilidad se incrementará progresivamente[30].

Leibniz tiene una clara conciencia metódica para distinguir entre una argumentación puramente racional y una ciencia de la experiencia. Sucede que la filosofía individualista monadológica leibniziana no concuerda con la silogística tradicional, el legado aristotélico de la racionalidad como generalidad que no atribuye racionalidad científica a los enunciados singulares. Pues aunque el aristotelismo comience en las cosas, construye en las *generalia* las bases lógicas para alcanzar el más alto conocimiento, la metafísica; es decir, tiene su acción en la *abstractio* de las cosas. Leibniz capta la paradoja de que analizar el individuo sea hacerle general, lo que equivale a alejarse de la realidad captable[31]. Va a intentar reconciliar las exigencias de la razón con el individuo, de modo que la ciencia del individuo deje de ser una contradicción terminológica. Es tarea urgente de la filosofía aceptar lo singular dentro de la teoría lógica, abrir la lógica a lo singular de modo que encuentre en ella su lugar y descubrir nuevos métodos para la fundamentación de las ciencias empíricas[32].

Leibniz, maestro en el empleo de técnicas formales del pensamiento simbólico, desarrolla un conjunto complementario de métodos y procedimientos racionales que contienen brillantes intuiciones para tratar las verdades de hecho de las ciencias experimentales, el ámbito no estrictamente demostrativo del saber. Para la adecuada descripción de las cosas de nuestro mundo es esencial la utilización de leyes de carácter relacional, al ser precisamente relacional –dotada de una estructura molecular– la naturaleza de los individuos. Lejos de la rigidez del silogismo, el *inesse,* la forma

---

[29] *Lettre de M.L. sur un principe general utile...*, GP III, 52.

[30] E. de Olaso, "Escepticismo e infinito", *Dianoia* XXXIII, nº 33, 1987, p. 229-31. Cf. "…adeoque infinita involvit, ideo nunquam perveniri potest ad perfectam demonstrationem, attamen semper magis magisque acceditur, ut differentia sit minor quavis data", GI § 74, A VI, 4 A, 763 ; "...les infinis ou les infiniment petits n'y signifient que les grandeurs qu'on peut prendre aussi grandes ou aussi petites que l'on voudra, pour montrer qu'une erreur est moindre que celle qu'on a assignée, c'est à dire qu'il n'y a aucune erreur…", *Théodicée* § 70, GP VI, 90 ; GI § 65, A VI, 4 A, 760.

[31] G. Aliberti, "Über das Individuationsprinzip: der junge Leibniz und die Auflösung der allgemeinen Substanz in der individuellen Substanz", en: *Nihil sine ratione: Mensch, Natur und Technik im Wirken von G.W. Leibniz,* VII Internationaler Leibniz-Kongreß. Hrsg. von Hans Poser, Hannover, 2001, pp. 9-10.

[32] E. Grosholz / E. Yakira, "Leibniz's science of the rational", *Studia Leibnitiana*: Sonderheft 26 (1998), p. 30.

sujeto-predicado, se entiende ahora como una relación más amplia que la de la parte y el todo. A partir de ella se organizan diversas maneras de composición entre términos homogéneos, es decir, que pueden sustituirse entre si *salva veritate*[33] estableciendo relaciones de coincidencia, lo que permite tratar una infinidad de combinaciones: de simetría, de conmutabilidad y transitividad[34]. La ciencia de las relaciones formales del pensamiento despliega estructuras topológicas a partir de las formas, la similitud, el orden y las disposiciones. Está introduciendo Leibniz la relación en el corazón de la lógica: deja de quedar prisionero de una lógica predicativa de nociones para pasar a desarrollar una lógica demostrativa de las relaciones, en donde no son los elementos comparables sino las estructuras. Es preciso componer una nueva lógica con una visión más amplia de forma que trascienda los límites de la lógica tradicional dando el paso desde la lógica conceptual aristotélica a la lógica de las relaciones, del concepto de relación entre elementos discretos al concepto de función entre variables, es decir, la asociación por correspondencia biunívoca de elementos que pertenecen a conjuntos diferentes. Se trata ahora de una lógica de lo parecido y lo desemejante, que progresa por proporciones equipolentes y conduce a fórmulas equivalentes, en donde las categorías aristotélicas de cualidad y cantidad, de espacio y tiempo quedan reducidas a relaciones. Lo esencial no es pues el contenido de los conceptos, como en la lógica tradicional, sino sus relaciones mutuas.

La *Combinatoria Specialis,* ciencia de las formas o cualidades, va a facilitar la transición a esa lógica de relaciones que Leibniz construirá hasta su vejez. Apunta a encontrar las continuidades que conectan las oposiciones, a fórmulas que faciliten la transición, a construir mediaciones en donde la armonía, concepto unificador central en su filosofía, implica la continuidad como instancia mediadora universal. Con el uso sistemático de la analogía[35], un esquema metodológico básico como fundamento de inducciones válidas, se desvía Leibniz del modo analítico de demostración. Metodológicamente anticartesiana, la sustitución combinatoria pasa a ser un importante aspecto epistemológico.

El marco de la lógica deductiva clásica tampoco permite la formalización de una teoría de la probabilidad. Sin embargo en la nueva lógica que Leibniz quiere desarrollar y que va más allá de una teoría abstracta del ámbito de lo necesario, se incluye la reducción de las inferencias probabilísticas a un cálculo, su genial idea de una matematización de la teoría de la probabilidad que supere el racionalismo dogmático de Descartes y Spinoza en cuanto a la valoración de lo probable[36]. El punto de arranque

---

33 "Eadem seu coincidentia sunt quorum alterutrum ubilibet potest substitui alteri salva veritate", GP VII, 236.

34 D. Berlioz, "Logique et métaphysique du meilleur des Mondes: le statut de l'*inesse*", *Studia Leibnitiana*: Sonderheft 21 (1992), pp. 171-72.

35 "...analogía, seu ipsarum similitudinum comparatio", *Characteristica verbalis*, 1679, AVI, 4 A, 336.

36 "Nam etiam probabilitates calculo ac demonstrationi subjiciuntur, cum aestimari semper possit, quodnam ex datis circunstantiis probabilius sit futurum", *De numeris characteristicis ad linguam universalem constituendam,* 1679, A VI, 4 A, 269. Cf. "J'ay dit plus d'une fois qu'il faudroit une *nouvelle espece de logique*, qui traiteroit des degrés de probabilité... une balance necessaire pour peser les apparences et pour former là dessus un jugement solide",

de la reflexión leibniziana sobre una lógica de la probabilidad fueron sus trabajos tempranos de carácter jurídico[37]. Efectivamente, la noción de probabilidad es esencial en la práctica jurídica, en concreto el concepto de *presunción* que admite alegaciones por anticipado pero con fundamento suficiente, en espera de una prueba de lo contrario. En el plan general de teoría probabilística que persigue nuestro autor, la lógica jurídica representa un caso especial[38].

Con su temprano escrito *De incerti aestimatione*, Leibniz es pionero en el desarrollo del cálculo de probabilidades como método probatorio[39]. Se trata de otro sentido de conocimiento: "la connoissance du vraisemblable" que define como el grado de posibilidad[40] y que introduce un cambio pragmático frente a la concepción lógico-metodológica, pues convenientemente desarrollada nos llevaría si bien no a la infalibilidad teórica, al menos a la infalibilidad práctica, apoyada en la coherencia y el éxito por los avances que se vayan produciendo. La ambiciosa heurística del proyecto leibniziano, en el deseo de fundamentar una lógica de la probabilidad o *Ars conjectandi*[41], sigue un procedimiento no cartesiano, al determinar lo cierto a partir de lo incierto. La variante probabilística como método racional se utiliza cuando no se puede proceder a la explicación integral de las causas posibles, sino únicamente a una

NE IV, 16, 9, A VI, 6, 466; "Et quant à la grandeur de la consequence et les degrés de probabilité nous manquons encore *de cette partie de la Logique*...", NE II, 21, 66, A VI, 6, 206.

37 *De conditionibus*, 1665, A VI, 1, 99-150; *Specimina Juris*, 1667-69, AVI, 1, 367-430.

38 "*...ut Mathematicos in necessariis, sic Jurisconsultos in contingentibus Logicam...*", *Ad stateram juris...,*C 211. Cf. "Et toute la *forme des procedures* en justice n'est autre chose en effet qu'une espece de *Logique*, appliquée aux questions de droit", NE IV, 16, 9, A VI, 6, 465.

39 A VI, 4 A, 91-101, septiembre 1678. Cf. "L'*opinion*, fondée dans le vraisemblable, merite peut etre aussi le nom de connoissance... je tiens que *la recherche de degrés de probabilité,* seroit très importante, et nous manque encor, et c'est un grand defaut de nos Logiques. Car lorsqu'on ne peut point decider absolument la question; on pourroit toujours determiner le degré de vraisemblance *ex datis*, et par consequent on peut juger raisonnablement quel parti est le plus apparent", NE IV, 2, 14, A VI, 6, 372; "Itaque, quando ex datis quaesitum non est determinatum aut exprimibile, tunc alterutrum hac analysi praestabimus, ut vel in *infinitum appropinquemus*, vel, quando conjecturis agendum est, demonstrativa saltem ratione *determinemus ipsum gradum probabilitatis*, qui ex datis haberi potest, sciamusque quomodo datae circumstantiae in rationes referri debeant, et quasi in bilancem, acceptis expensisque similem, redigi queant, ut quod maxime rationi consentaneum sit eligamus", GP VII, 201; "Sed doctrina de verisimilitudine nondum a quoquam pro dignitate tractata est... Hanc Logicae partem inter desiderata colloco...", *A Koch*, GP VII, 477; "Hoc autem recte fit, cum ex plurimus eligimus verisimilius…", C 496. En época de Leibniz es Huygens quien inicia el cálculo de probabilidades con su obra *De ratiociniis in ludo aleae* (1657). Mientras que Jacques Bernoulli con su *Ars conjectandi* (1713) escribe el principal tratado sobre la probabilidad hasta la *Théorie analytique des probabilités* (1812) de Laplace.

40 "*Probabilitas* est gradus possibilitatis", *De incerti aestimatione,* septembre 1678, A VI, 4 A, 94. Cf. "Sunt quidem in inquiriendo gradus", *A Bierling,* 12 agosto 1711, GP VII, 500; "Il y a donc une science sur les matieres les plus incertaines, qui fait connoistre demonstrativement les degrés de l'apparence et de l'incertitude…", *Nouvelles ouvertures,* 1686, A VI, 4 A, 689.

41 *De arte inveniendi Theoremata*, 7 septiembre 1674, A VI, 3, 426.

aproximación. El cálculo de probabilidades sustituye, a nuestra medida humana y limitada, al cálculo infinito divino. No podremos como Dios conocer el edificio del saber *a priori*, pero podremos reconstruirlo progresivamente *a posteriori.* El *Ars inveniendi* o arte de descubrir nuevos conocimientos para la ciencia mediante la combinatoria es un *ars cryptographica,* por el que la creación divina se presenta como un texto a descifrar.

Vemos cómo Leibniz redefine el esquema metodológico universal constituido por el binomio análisis/síntesis y lo completa con otros esquemas metodológicos. Pero hay que tener en cuenta que en su ideal de unificación la *Scientia Generalis* cubre lo pensable en general y el conjunto de todos los principios de las demás ciencias[42], es decir, abarca tanto el *ars inveniendi* como el *ars judicandi*[43]. Ambos artes se distinguen en sus fines específicos y en sus métodos: el *ars inveniendi* apunta a formular nuevas teorías y obtener nuevos resultados, mientras que el *ars judicandi* verifica las teorías determinando si son aceptables las nuevas verdades. No se oponen entre sí, sino que están en estrecha relación de complementariedad, por lo que dice de ellos que "ces deux arts ne different pas tant qu'on croit"[44].

En *De rationibus motus*, uno de los más interesantes escritos del período de juventud de Leibniz, utiliza la metáfora de comparar las ciencias con un río que desemboca en un océano de múltiples usos y métodos[45]. La variedad de perspectivas se extiende también al nivel metodológico. En la búsqueda de una organización sistemática de los saberes, de las condiciones de inteligibilidad y de racionalidad de lo real, la *Scientia Generalis* se presenta como una colección ordenada de demostraciones ciertas y probables, estrategias múltiples probatorias tanto *a priori* como *a posteriori*. En definitiva, la filosofía leibniziana reúne una complejidad de métodos y técnicas para la obtención y validación del saber, una diversidad de modelos metodológicos en los que Leibniz despliega su enorme capacidad creativa, pues como él mismo dice "Si j'estois aussi capable d'achever des Methodes, que je suis disposé à en projetter, nous irions sans doute bien loin..."[46].

---

[42] "Scientiam generalem intelligo, quae caeterarum omnium principio continet…", D*efinitio brevis Scientiae Generalis*, 1683-85 ?, A VI, 4 A, 532.

[43] "…la SCIENCE GENERALE qui doit donner… la Methode de juger et d'inventer…", *Nouvelles ouvertures*, 1686, A VI, 4 A, 691.

[44] *Discours touchant la methode…*, 1688-90 ?, A VI, 4 A, 962. Cf. "...conjungi debent inventionis lux, et demonstrandi rigor...", *Consilium de Encyclopaedia…,* 15 junio 1679, A VI, 4 A, 342; "Unter der *Logick* oder Denckunst verstehe ich die Kunst den verstand zu gebrauchen, also nicht allein was fürgestellet zu beuertheilen, sondern auch was verborgen zu erfinden", *A Wagner*, GP VII, 516; M. Dascal, "Leibniz y las tecnologías cognitivas", en: *Actas del Congreso Internacional Ciencia, Tecnología y Bien Común: La Actualidad de Leibniz,* op. cit., p. 177.

[45] "…nam fontes artium, ut ariditate quadam et simplicitate delicatis displicere solent, ita decursu perpetuo in uberrima scientiarum flumina, denique quoddam velut mare usus ac praxeos excrescunt", *De rationibus motus*, 1669, A VI, 2, 160.

[46] *A l'Hospital*, 28 abril 1693, A III, 5, 542.

Arquitecturas emocionales en inteligencia artificial

M.G. Bedia[1], J.M.Corchado[2] y J. Ostalé[3]

[1] Departamento de Informática, Universidad Carlos III de Madrid
Av. de la Universidad, 30, 28911, Leganés (Madrid), España
mgbedia@inf.uc3m.es
[2] Departamento de Informática y Automática, Universidad de Salamanca
Plaza de la Merced s/n, 37008, Salamanca, España
corchado@usal.es
[3] Departamento de Filosofía, Lógica y Filosofía de la Ciencia, Universidad de Salamanca
Campus Miguel de Unamuno s/n, 37007, Salamanca, España
ostale@usal.es

Resumen. Las emociones pueden ser explicadas de diferentes modos dependiendo de las cuestiones que nos planteemos. En Inteligencia Artificial el aspecto que interesa del estudio de las emociones, es el de buscar qué tipos de requisitos estructurales son satisfechos por los estados emocionales y qué mecanismos funcionales subyacen a los procesos emotivos, con el fin de reproducirlos en arquitecturas artificiales. En este artículo se hace una revisión de las arquitecturas emocionales más importantes que permite explicar el error de muchas de las estrategias de implementación de modelos emocionales: subyace el carácter dualista emoción-razonamiento que se ha asumido tradicionalmente. Los trabajos más modernos en neurofisiología de las emociones han llevado a los investigadores en Inteligencia Artificial a mostrar un aumento de interés por el diseño de sistemas emocionales y a plantearse nuevas arquitecturas sin proponerse, en realidad ni cuestionarse, si los sistemas artificiales pueden tener realmente emociones como las de los humanos: las emociones serán reproducidas exclusivamente desde una perspectiva funcional.

1 Introducción

Existen fundamentalmente tres razones por las que debería interesarnos el estudio y la implementación de arquitecturas emocionales en Inteligencia Artificial. En primer lugar, por la existencia de problemas en Inteligencia Artificial para los que la aplicación de métodos clásicos no funcionan: necesitamos herramientas de diseño más efectivas y existen teorías emocionales que pueden inspirar su desarrollo. [Damasio, 1996.] En segundo lugar, existen algunas teorías sobre las emociones en Neurología que necesitan aún soporte empírico. La implementación en modelos artificiales junto a experimentos de simulación pueden ser útiles para contrastar este tipo de teorías sobre el comportamiento emocional de animales y humanos. Y, por ultimo, si asumimos que el objetivo fundamental de la Inteligencia Artificial es el desarrollo de sistemas artificiales que actúen de manera similar a los humanos, y las emociones tienen funciones biológicas importantes, como así parecen asegurarlo numerosos estudios

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 186-193, 2006.

[LeDoux, 2000], entonces el diseño de sistemas emocionales artificiales incorporará nuevas funcionalidades a los artefactos que permitirán mejores competencias y mayor autonomía.

2 Un repaso de las emociones desde el enfoque de la IA

En este artículo nos proponemos presentar las distintas concepciones que se han ido planteando sobre las emociones a lo largo de la historia, y también las estrategias emocionales más representativas que, en el contexto de la IA, se han desarrollado desde la década de los setenta hasta la actualidad. A continuación, relacionaremos las hipótesis implícitas en el diseño de tales sistemas con los principios de diferentes teorías emocionales en áreas como la Psicología o Neurociencia. Por último, presentaremos las razones por las que tales estrategias no han obtenido el éxito esperado y propondremos una posible vía de solución.

2.1 Teorías emocionales

Existen fundamentalmente tres clases de teorías para explicar el comportamiento de las emociones:

- **Teorías No cognitivas (o anti-cognitivas)**
  Tradicionalmente se han considerado las emociones como estados mentales, conscientes y por tanto identificables mediante el lenguaje. Desde esta perspectiva, la mente sería la combinación, por un lado de razones y por otro de emociones, ámbitos separados y con estatus diferentes. Para este tipo de explicaciones, las emociones serían esencialmente un modo de respuesta de caracter reactivo, codificadas en nuestros genes, que funcionarían como alarmas ante la percepción de situaciones peligrosas para nuestra supervivencia. Este enfoque no es satisfactorio por diversas razones [Izard, 1977; Zajonc, 1980] y no resuelve algunas preguntas sobre las emociones y la inteligencia.
- **Teorías Cognitivas**
  El primer trabajo donde se considera que las emociones no constituyen un módulo al margen de la cognición fue [James, 1884]. Por primera vez se plantea una teoría que interpreta las emociones como la toma de conciencia de las reacciones viscerales (y no las reacciones mismas). Este carácter cognitivo de las emociones permite entender emociones complejas que no podían explicarse con el modelo anti-cognitivo (como emociones nostálgicas acerca de algo que ocurrió) [Lazarus, 1984] pero no resuelve algunos aspectos que sí explicaban las teorías anteriores.
- **Teorías interactivas**
  Con este tipo de teorías se pretende encontrar un modelo que integre los resultados de ambas teorías previas. En este caso, las emociones son una entidad producto de dos aspectos hasta el momento separados: activación visceral (anticognitivista) y valoración cognitiva de la alteración (cognitivista) [Schachhter &

Singer, 1962, Arnold, 1960]. El modelo interactivo más completo (y satisfactorio) es el modelo de Frijda [Frijda, 1986]. Permite integrar y explicar emociones primarias (reactivas), secundarias (evaluadas cognitivamente) y meta-emociones de carácter contrafactual (no existen motivos que las provocan).

2.2 Arquitecturas emocionales

Vamos a mostrar algunas arquitecturas emocionales que podemos mostrar como ejemplo práctico de las teorías antes presentadas.

Arquitectura No cognitiva [Tomkins, 1984]
El comportamiento emocional en este tipo de sistemas no está relacionado con ningún tipo de sistema cognitivo-deliberativo. La arquitectura establece relaciones entre estímulos y respuestas pre-establecidas mediante un principio de excitación basado en la "densidad de disparos" recibidos (intensidad del estímulo por tiempo). En esta arquitectura, las emociones que pueden generarse cumplen las siguientes características: (1) las emociones como un tipo de reacciones automáticas, (2) la emoción que se manifiesta es la que responde a mayor estimulación de la siguiente manera:

- Si la estimulación aumenta se representan estados como miedo o interés
- Si la estimulación decrece se tienen estados como alegría
- Si la estimulación se nivela tenemos estados como la angustia o el enojo.

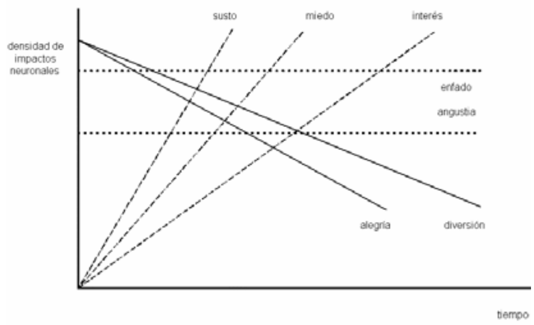


Fig. 1. Modelo de Tomkins

Arquitectura cognitiva [Simon, 1982]
En este modelo, las emociones no se entienden como reacciones automáticas sino más bien como elementos evaluadores del entorno con la capacidad para interrumpir el proceso deliberativo en curso y forzar al sistema para que centre su atención y sus recursos en una nueva situación.

Las únicas emociones que pueden representarse son aquellas que se pueden entender como mecanismos de atención urgente. El sistema básicamente está formado por (1) un módulo central de gestión de objetivos de tipo serial (los nuevos estímulos se ponen a la cola y esperan turno para ser evaluados organizándose en forma de jerarquía de objetivos) y (2) Un módulo emocional representado por un sistema de vigilancia con capacidad para interrumpir el procesamiento del módulo central si observa contingencias que requieren atención urgente.

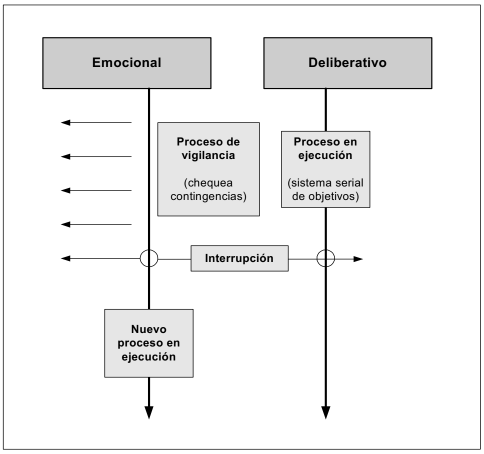


Fig.2. Modelo de Simon

Arquitectura interactiva [Sloman, 1987]

Este modelo es más complejo aunque se deriva del anterior. Presenta emociones del tipo del modelo de Simon pero incorpora:

1. Un mecanismo de filtro: las actividades en curso usan recursos, tanto cognitivos como físicos, que son limitados. Por tanto, es necesario un filtro para que estén protegidos y abiertos a la interrupción.
2. De umbral variable: La variabilidad del umbral permite al nivel de protección ser dependiente del contexto.

El modelo de Sloman además de tener la capacidad de interrumpir los objetivos actuales del sistema puede representar la disposición a la interrupción sin llegar a hacerlo.

En este modelo, mucho más rico, podemos reproducir comportamiento emocional con las características de las emociones de Simon, es decir, *emociones como mecanismos de atención urgente que interrumpen el procesamiento actual,* pero además con los siguientes propiedades: (1) Emociones como mecanismos disposicionales, (2) Emociones como mecanismo graduales y (3) Emociones como disposiciones que pueden coexistir.

Arquitectura basada en el modelo de Frijda [Frijda, 1986]

Siguiendo el modelo de Frijda se desarrolla la propuesta más completa hasta el momento. Se estructura a partir del modelo de Sloman añadiendo un canal para el flujo de información continuo y bidireccional entre proceso de atención, vigilancia y filtro. Bajo estas condiciones podemos tener flujos de información, a los que podemos asociar significado emocional, dadas las propiedades que presentan:

- Pueden interrumpir el proceso de atención si detectan situaciones de emergencia (característica del modelo de Simon)
- Pueden representarse como disposiciones a la acción, coexistir varias y presentan un carácter gradual (característica del modelo Sloman)
- Su manifestación depende de una articulación entre percepción, contexto y experiencia (característica del modelo de Frijda)

Una de las arquitecturas emocionales de mayor éxito, TABASCO (Tractable Appraisal-Based Architecture for Situated Cognizers) [Petta, 2002], sigue en su diseño un modelo emocional interactivo basado en la teoría de Frijda.

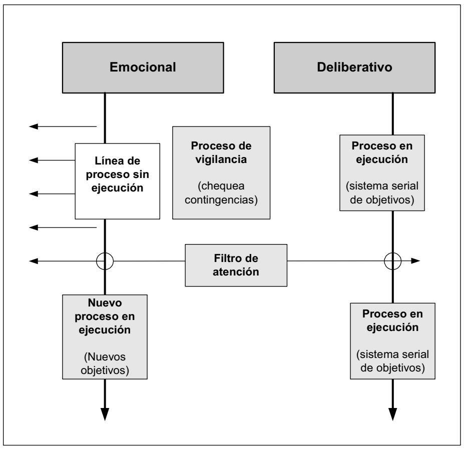


Fig.3. Modelo de Sloman

3 Problemas y limitaciones

En el conjunto de todas las arquitecturas encontramos una serie de elementos que se repiten: Las emociones son prefijadas y formalizadas como conceptos, y actúan esencialmente como perturbaciones, esto es, interrumpen el proceso de deliberación en curso. Estas condiciones definen un comportamiento prefijado para las distintas emociones y una serie de características típicas:

- *Contenido representacional:* Emociones formalizadas como conceptos.
- *Contenido no-representacional*: Emociones modeladas como señales de alarma.
- *Capacidad reactiva de tiempo real*: Mecanismo de vigilancia e interrupción.
- *Disposiciones sin interrupción*: Emociones como motivaciones a partir de un mecanismo de filtro variable.

Sin embargo, nos gustaría que se presentasen otras características que también son rasgos esenciales de las emociones como, por ejemplo, que (i) generasen estructuras de *sentimiento*, es decir, un mecanismo de valoración emocional, que (ii) generasen procesos de adaptación emocional (*aprendizaje emocional*).y (iii) que favoreciesen procesos de decisión. Según teorías neurofisiológicas la falta de emociones en sujetos inteligentes puede constituir una fuente importante de conducta irracional. Se ha comprobado [Damasio, 1996] que pacientes con daños en el lóbulo central, volviéndose emocionalmente planos, pierden su capacidad para tomar decisiones racionales.

4 Posibles soluciones

Para reproducir artificialmente un comportamiento emocional más complejo y más ajustado al que observamos en la realidad, sugerimos adentrarnos en *las teorías Neurofisiológicas* acerca de las emociones. En estas teorías se defiende un enfoque completamente diferente al de los modelos de emociones basadas en conceptos.

Dualismo vs. Funcionalismo
Mientras que los modelos emocionales basados en conceptos presentan habitualmente una visión dualista entre la cognición y las emociones, la Neurofisiología no establece diferencias entre ambos: los dos son casos de procesamiento informacional. Esta posición de cara al estudio implica que no nos interesarán explicaciones causales sobre los mecanismos que las generan (¿*Cómo ocurren*?), tampoco nos preocuparemos por teorías evolutivas que expliquen su relación con las funciones mentales (¿*Por qué ocurren así*?) ni cuestiones acerca del carácter innato, aprendido, o las modificaciones a lo largo del desarrollo del individuo (¿*Cuándo ocurren*?).

Nuestro interés se centra en entender "*para qué sirven"*, es decir, cómo las emociones hacen a los organismos sobrevivir o adaptarse. De otra forma, *"cuál es su función"*.

4.2 Perturbaciones vs. Mecanismos de apoyo a la razón
Los trabajos de [Damasio, 1996; De Sousa, 1987] defienden que las emociones *nos ayudan a tomar decisiones,* pues, por un lado, deshacen el empate en los casos de *indiferencia,* y por otro lado*,* constituyen un criterio de selección en casos de *indeterminación*, al hacer posible que nos centremos en los rasgos más destacados de la situación presente. Por tanto, pasamos de entender las emociones como "interrupciones" a considerarlas como sistemas de toma de decisiones que, junto a mecanismos deliberativos y en ciertas ocasiones, mejoran nuestra conducta.

5 Conclusiones y trabajo futuro

El trabajo que pretendemos llevar adelante, tras el estudio realizado, es encontrar modelos de arquitecturas emocionales con las siguientes características:

- Neutras (sin emociones pre-definidas).
- Emociones como procesos funcionales.
- Emociones como mecanismos de supervivencia (mejora del comportamiento del sistema).
- Emociones como mecanismos en procesos de toma de decisión (planificación) complementarios a la deliberación.

Y, por ultimo, todas estas propiedades deberían poder considerarse como elementos de un modulo emocional que no se asocie a arquitecturas ya desarrolladas sino que se integre en la arquitectura desde un punto de vista constructivo (figura 4).

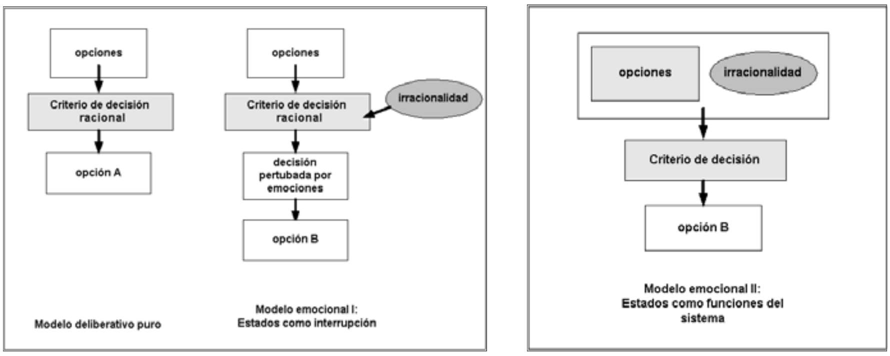


Fig.4. Cambio en el diseño de arquitecturas emocionales: (a) arquitecturas clásicas, (b) propuesta de arquitectura emocional "contructiva"

Referencias

[Arnold, 1960] Arnold, M. (1960). Emotions and Personality. Nueva York. Columbia University Press.

[Damasio, 1996] Damasio, A. (1996). Descartes' error. Drakontos, Crítica.

[De Sousa, 1987] De Sousa, R. (1987). The Rationality of Emotion. Cambridge, Mass., MIT Press.

[Frijda, 1986] Frijda, N. (1986). The emotions. Sudies in Emotion and Social Interactions. Cambridge University Press, Cambridge, UK.

[Izard, 1977] Izard, C. E. (1977). Human Emotions. Nueva York. Plenum Press.

[James, 1884] James, W. (1884). What is emotion?, Mind, 1884, pp. 59

[Lazarus, 1991] Lazarus, R.S. (1991). Progress on a Cognitive-Motivational-Relational theory of emotion. American Psychologist, No. 46, pp. 819-834.

[LeDoux, 2000] LeDoux, J. (2000). The Emotional Brain, Nueva York, 2000

[Petta, 2002] Petta, P. (2002). The role of emotions in tractable architectures for situazed cognizers. In Trappl, R., Petta, P. and Payr, S. (Eds.). Emotions in Humans and Artifacts. Cambridge, M.A., MIT Press.

[Schachhter, 1962] Schachhter, S. and Singer, J. (1962). Cognitive, social and physiological determinants of emotional state. Psychological Review, 59, pp. 379-399.

[Simon, 1982] Simon, H.A. (1982). A□ect and Cognition: Comments. In Clark, M.S. and Fiske, S.T. (Eds.). The Seventeenth Annual Carnegie Symposium on Cognition: A□ect and Cognition. London: Lawrence Erlbaum Associates, pp. 333-342.

[Sloman, 1987] Sloman, A. (1987). Motives mechanisms and emotions. Cognition and Emotion, 1(3), pp. 217-234. Reprinted in Boden, M.A. (Ed.). The philosophy of Arti□cial Intelligence, OUP.

[Tomkins, 1984] Tomkins, S. (1984). A□ect Theory. In Scherer, K. and Ekman, P. (Eds.). Approaches to Emotion. Hillsdale, NJ: Lawrence Erlbaum Associates.

[Zajonc, 1980] Zajonc, R. B. (1980). Feeling and thinking: preferences need no inferences. American Psychologist, 35, pp. 151-175.

[Nolfi and Floreano, 2000] Nolfi, S. and Floreano, D. (2000). Evolutionary Robotics. The Biology, Intelligence, and Technology of Self-organizing Machines. MIT Press.

[Norton, 1995]. Norton, A. (1995). Dynamics: An Introduction. In Port, R. and Van Gelder, T. (Ed.). Mind as Motion: Explorations in the Dynamics of Cognition. Cambridge University

Press.
[Rabinovich, 2000]. Rabinovich, M. I., Varona, P. & Abarbanel, H. D. I. Nonlinear cooperative dynamics of living neurons. Int. J. .Bifurcation Chaos 10 (5), 913-933 (2000)
[Rabinovich et al. 2001] Rabinovich, M.I., et al., Dynamical encoding by networks of competing neuron groups: Winnerless competition. Physical Review Letters, 2001. 87 (6), (2001)
[Schoener et al., 1995] Sch□oner, G., Dose, M., and Engels, C. (1995). Dynamics of behaviour: theory and applications for autonomous robot architectures. Robotics and Autonomous Systems, 16:213-245.
[Schoener et al., 1998] Schoener, G., Dijkstra, T., and Jeka, J. (1998). Action-perception patterns emerge from coupling and adaptation. Ecological Psychology, 10:323-346.
[Townsend, J. Busemeyer,J., 1995] Townsend, J. Busemeyer,J.Dynamics representation of decision-making, in: R.F. Port, T. Van Gelder (Eds.), Exploration in the Dynamics of Cognition: Mind as Motion, MIT Press, 1995, pp. 101–120.
[Verschure, 2003] Verschure, P., Voegtlin, T. & Douglas, R. J. Environmentally mediated synergy between perception and behaviour in mobile robots. Nature 425, 620-624 (2003)
[Woergoetter and Porr, 2005] Woergoetter, F. and Porr, B. (2005). Temporal sequence learning, prediction and control. Neural Computation, 17:245-319.
[Wolpert et al., 1995] Wolpert, D., Ghahramani, Z., and Jordan, M. (1995). An internal model for sensorimotor integration. Science, 26:1880-1882.
[Wolpert and Ghahramani, 2000] Wolpert, D. and Ghahramani, Z. (2000). Computational principles of movement neuroscience. Nature Neuroscience, 3:1212-1217.

Una perspectiva naturalizada del concepto de información en el sistema nervioso

Xabier Barandiaran[1] y Álvaro Moreno[2]

[1] Universidad del País Vasco - Euskal Herriko Unibertsitatea,
Post Box 1249, 20080 Donostia - San Sebastian, Gipuzkoa
xabier@barandiaran.net,
http://ehu.es/ias-research/barandiaran
[2] ylpmobea@sf.ehu.es,
http://ehu.es/ias-research/moreno

Resumen En este trabajo planteamos una revisión crítica del concepto de información neuronal, tal como ha sido heredado de la inteligencia artificial clásica dentro del paradigma dominante en neurociencias. A partir de una concepción del organismo como sistema autónomo, reconstruimos desde una perspectiva naturalizadota un nuevo concepto de información neuronal. La información aparece como estructura dinámica desacoplada de los procesos metabólicos enraizada en las interacciones sensomotrices de un organismo y recurrentemente seleccionada por éste en virtud de su contribución a su automantenimiento del sistema.

Palabras clave: información, sistema nervioso, naturalización, autoorganización.

1. Los problemas de la concepción heredada de la información en neurociencia

El concepto de información es, sin duda, uno de los pilares fundamentales del la IA y de las ciencias de la computación. A su vez, el feed-back conceptual y metodológico entre la IA y la neurociencia cognitiva ha sido contínuo desde el surgimiento de la cibernética y el desarrollo de la IA a partir de los años 50 (especialmente en lo que se refiere al concepto de información). De hecho, la literatura neurocientífica está cargada de términos que superan el marco descriptivo molecular o bioquímico, introduciendo conceptos como el de *señal, código, información o contenido* desde que en los años 20 y 30 comenzara por primera vez a hablarse de los impulsos nerviosos como vehículos de mensajes [1]. Sin embargo, no sería hasta el desarrollo del concepto de información de Shanon y, sobre todo, del uso generalizado del concepto de representación e información en el campo de la inteligencia artificial (a finales de los años 50) que el uso del término información se extendiera en la literatura neurocientífica. 50 años después, algunas de

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 194-206, 2006.

las características más fundamentales del concepto de información (la semántica entendida como intencionalidad) generan problemas que siguen sin poder resolverse dentro del paradigma computacionalista y representacionalista, comunes tanto en neurociencia como en la IA clásica.

La forma más común en neurociencias de referirse a la información es una forma metodológica o descriptiva, mediante la cual un investigador decide usar el término información para nombrar ciertas regularidades en los fenómenos que observa. El uso de los términos información, código, contenido, significado, etc. en neurociencia teórica está asociado a la correlación entre un estímulo dado y un potencial de acción (o un conjunto de potenciales de acción) en un área específica del sistema nervioso (SN). Esta correlación causal está generalmente expresada como probabilidad condicional de la ocurrencia de un estímulo, dada una medida de impulsos o una combinación de ellos [2], [3]. Esta perspectiva nos permite como mucho predecir (*a posteriori* —después de toda una serie de experimentos) la probabilidad de disparo de una neurona en relación a un estímulo determinado. Hasta aquí la información cumple un papel descriptivo en forma de probabilidad condicional. Sin embargo, implícita o explícitamente, el uso del término viene generalmente asociado a un valor semántico de tipo referencial: el del objeto o propiedades del objeto que hace de estímulo y aparece por tanto correlacionado con mayor o menor probabilidad con la actividad de diversas áreas del SN.

Este concepto de información (y toda la investigación neurocientífica que se realiza a partir de ella) se enfrenta, sin embargo, a un problema fundamental. La medida de correlación condicional estímulo-actividad nerviosa en ningún caso *explica* el funcionamiento del SN *en relación a esas correlaciones*. De hecho, la correlación (así explicada, como probabilidad condicional) no es nunca accesible al propio SN sino sólo al observador que accede al estímulo y sus consecuencias dinámicas de forma separada, de tal manera que el proceso resulta informacional para el observador pero no necesariamente para el sistema bajo estudio. En otras palabras: si el contenido semántico está asegurado por la correlación condicional pero ésta es inaccesible para el sistema cognitivo …¿en qué sentido supone la probabilidad condicional una explicación semántica de la conducta del sujeto? ¿en virtud de qué se crean y se transforman (corrigen, desechan, modifican) esas correlaciones? ¿cuales son los mecanismo internos que retienen las correlaciones y en virtud de qué (ya que la comparación entre estados del mundo y actividad nerviosa le está, en principio, vedada al SN)? Estas preguntas sugieren que tratemos de dar cuenta del contenido semántico de un proceso informacional desde la perspectiva del organismo, en virtud del modo particular en el que se inserta en su organización conductual, y no desde la perspectiva de un observador externo cuyo acceso privilegiado al origen del estímulo y a la actividad nerviosa permite establecer una correlación entre ambos. Este uso del término información está generalmente asociado a la presuposición (implicita o explícita) de un homúnculo que hace las veces de intérprete del contenido semántico. Lo que permite afirmaciones como la siguiente: "(…) we can ask how the homunculus should best use the spike train data to make a decision about which stimulus

in fact ocurred." ([3], página 14). Sin embargo, no se trata tanto de inferir "la realidad" a partir de un conjunto de señales alojadas en alguna parte del SN, sino de explicar cómo se integran las señales informacionales en la producción de conducta, y en virtud de qué aparecen, para el organismo, asociadas a una intencionalidad o semántica acerca del mundo y devienen, por tanto, informacionales.

Por estas razones, esta concepción de la información neuronal nos parece insostenible. Resulta que, en la explicación de un fenómeno natural, el investigador tiene que echar mano de un concepto (el de información) que no puede reconstruir naturalizadamente (e.d. haciendo depender el valor *explicativo* de la semántica informacional en un observador externo que no es parte del fenómeno investigado). La perspectiva naturalizadora opta, en cambio, por una concepción ontológica en la que el término información se refiere a un tipo de causalidad específica en el SN, y que además permita dar cuenta del contenido semántico de un proceso informacional desde la organización misma del agente cognitivo (y, en concreto, de los organismos vivos). Se trata de especificar cómo un proceso físico o biológico deviene informacional no para un observador externo al que se le presuponen ya capacidades cognitivas, sino *en el propio marco* del sistema natural en el que ese proceso tiene lugar, jugando un papel causal específico al constituir una nueva forma de organización en dicho sistema. La cognición y la información con propiedades semánticas son fenómenos emergentes. Minutos después del *big-bang* no existía el fenómeno cognitivo en el universo (no había nada de lo que pudiéramos predicar capacidades cognitivas). Hoy, en cambio, atribuimos capacidades cognitivas a diversos sistemas que nos rodean, especialmente a nosotros mismos. ¿Cómo ocurrió esta transición en la historia del universo? ¿Qué tipo de organización característica de la materia lo posibilita? ¿Qué papel juega en ella la información? La respuesta a estas preguntas exige un breve recorrido por el origen y organización de la vida y, sobre todo, del SN.

2. La aparición del sistema nervioso en la organización viviente

Todo ser vivo individual es un sistema *autónomo* que desarrolla interacciones sobre su entorno como parte sustancial de su propio proceso de automantenimiento [4]. Esta capacidad de ejercer interacciones, sostenidas por su organización interna, lo constituye como un *agente*. Con la excepción de algún tipo de bacteria que vive en un entorno muy homogéneo y estable, todos los seres vivos tienen mecanismos internos que compensan en tiempo somático, dentro de ciertos márgenes, las diferentes condiciones de su entorno: es decir, son agentes *adaptativos*. Los organismos tienen la capacidad de detectar aquellas modificaciones en su entorno que son relevantes para su mantenimiento, y desencadenar ciertos procesos internos y externos que contribuyen a restablecer y mejorar las condiciones de su mantenimiento.

En los unicelulares y en las plantas la agencialidad adaptativa está soportada por la organización metabólica. Sin embargo, la organización metabólica no

permite soportar un rápido y versátil abanico de respuestas para los organismos multicelulares cuya forma de vida está basada en el movimiento. Sabemos que el problema se resolvió con la aparición del SN. Probablemente este hecho fue resultado de la combinación de dos factores: 1) la aparición (como consecuencia de un proceso de diferenciación celular) de un nuevo tipo de células (las neuronas) capaces de conectar de manera plástica, rápida y (metabólicamente hablando) económica, superficies sensoras y motoras; y 2) la aparición de algún tipo de metazoo cuyo plan corporal permitiera reclutar a estas células para sostener formas de agencialidad sensomotriz muy sencillas. La potenciación de la conducta motriz que, presumiblemente, debió de producirse sería la base de un proceso evolutivo en el que se seleccionarían aquellos animales que poseyeran redes de neuronas interconectadas, actuando como soporte de comportamientos sensomotrices funcionales.

Pero, además de permitir el desarrollo de la motilidad en organismos multicelulares, la aparición del SN ha abierto el paso a nuevas (cualitativamente diferentes) formas de interacción adaptativa, muchísimo más complejas que todas las demás. Ya en los primeros estadios evolutivos del SN aparecen formas rudimentarias de aprendizaje, categorización y memoria [5]. Esta potencialidad de soportar complejidad agencial reside en la capacidad que tiene el SN de generar un dominio propio con un enorme número de configuraciones, cuya dinámica concreta no puede ser especificada por la organización básica del organismo.

3. Estructura y función del sistema nervioso

Las neuronas presentan una serie de características propias ausentes en las células que forman otros tejidos o sistemas dentro del organismo[3]. La interacción entre la dinámica básica generada por los potenciales de acción y otros procesos de segundo orden, como su sincronización a diversas escalas temporales y espaciales, permite generar (con relativamente pocas neuronas) una gran complejidad dinámica. Estas características permiten una conectividad recursiva y una velocidad de interacción *selectiva* interneuronal en frecuencias de interacción muy superiores a las que rigen cualquier otro proceso de control en el organismo. Esto dota a la red interneuronal de unas características especiales en el conjunto del organismo: ningún otro sistema intercelular tiene la capacidad, ni de lejos, que tiene el SN de correlacionar funcionalmente tantas unidades y, al

[3] Como son: a) la capacidad para el bombeo activo de iones que permite la propagación (sin disipación, e.d. sin perdida de variabilidad) del cambio del potencial de membrana a lo largo del cuerpo celular y sus extremidades, b) conectividad conductiva con otras neuronas a través de superficies sinápticas con receptores químicos que permiten la regeneración (variable) del cambio de potencial de membrana de las neuronas vecinas, c) ramificaciones a largas distancias del cuerpo celular (desde 0.1mm hasta 3 metros) d) no-linealidad del potencial de respuesta y e) plasticidad (variabilidad) a múltiples escalas temporales y biofísicas en relación a las respuestas a los cambios de potencial y sinápticos (crecimiento dendrítico, variación de concentraciones de neurotransmisores en las sinapsis, cambios moleculares, etc.)

mismo tiempo, de modificar selectivamente los estados de dichas unidades tan rápidamente. *La particularidad del SN es, por tanto, su capacidad de generar una variedad enorme de estados (configuraciones) por unidad de tiempo y de coordinar un número inmenso de transformaciones de estado paralelamente.* Características todas ellas (como veremos) propicias para la generación de procesos autoorganizativos de regulación. Dicho de otro modo: las neuronas son un tipo extremadamente peculiar de células que se han seleccionado por su capacidad para mantener una compleja, plástica y rápida dinámica de interacciones entre sí, minimizando la interferencia con procesos metabólicos locales. Y con el resto de órganos y sistemas biológicos de carácter metabólico (sistema circulatorio, respiratorio, digestivo, etc.).

Este último factor (la minimización de interferencias con procesos metabólicos locales) va a ser de fundamental importancia, ya que va a definir el aspecto *desacoplado* del SN: e.d. su dinámica va a quedar subdeterminada por los procesos puramente metabólicos (autoconstructivos) de la estructura biológica que lo sustenta (el conjunto de la infraestructura neuronal). El desacoplamiento jerárquico del SN apunta a una doble condición del mismo: está localmente construido por procesos metabólicos (aspecto jerárquico: el SN emerge de procesos metabólicos) y, sin embargo, éstos no son suficientes para explicitar completamente su dinámica funcional (aspecto desacoplado) [6]. Otro factor que contribuye a esta subdeterminación de la dinámica neuronal por constricciones metabólicas es su acoplamiento adicional con su entorno sensomotor a través de la interfaz del cuerpo. Su corporización (*embodiment*) y su "estar situado" (*situatedness*) hacen que la dinámica del SN responda también a factores interactivos inscritos en ciclos sensomotores.

El SN presenta además una dinámica interna reticular, cohesiva y recurrente capaz de mantener patrones invariantes frente a perturbaciones internas y externas, de tal manera que su complejidad dinámica interna es mayor que la dinámica interactiva que establece con su entorno. Esta asimetría de complejidad hace que su dinámica aparezca en buena medida como *autodeterminada*. Si para un momento dado quisiéramos predecir la evolución de la dinámica neuronal, el conocimiento del estado metabólico de sus componentes aislados no nos serviría de gran cosa. Por el contrario, tendríamos que atender al contexto corporal y ambiental en el que se encuentra el organismo y, sobre todo, a la propia dinámica interna que es recursivamente generada atendiendo a sus propiedades holistas de red, más allá de las interacciones celulares locales.

Podemos ahora responder apropiadamente a la cuestión de la *función* del SN en el conjunto del organismo: el control de los ciclos de interacción (de la clausura interactiva necesaria para el automantenimiento del organismo) a través de una red desacoplada de la dinámica metabólica, pero inserta en una interfaz corporal (sistemas musculo-esquelético y perceptivo). Por un lado, podemos observar el flujo material y termodinámico que se realiza entre la ingestión y digestión de alimentos y oxígeno, su secreción y la disipación de calor en el cuerpo, junto a toda la maquinaria autoconstructiva del organismo y la infraestructura necesaria para su mantenimiento (sistema circulatorio, digestivo, respiratorio, etc.).

Por otro lado, observamos un ciclo sensomotor de conducta adaptativa regido por una infraestructura (sistema sensomotor y SN) dinámicamente desacoplada de los procesos metabólicos (aunque sustentada en ellos). La funcionalidad del SN viene dada por la forma en la que se conectan ambos procesos a un nivel superior. Esta relación funcional exige que la dinámica interactiva controlada por el SN satisfaga las condiciones de automantenimiento del organismo, más concretamente, debe satisfacer las condiciones de contorno del flujo termodinámico biológico. Desde el punto de vista del funcionamiento dinámico del SN, la satisfacción de estas condiciones de contorno del flujo termodinámico aparecen como lo que Ashby [7] denominó *variables esenciales*: un conjunto de variables (temperatura, input nutricional, etc.) que deben ser mantenidas dentro de unos límites de viabilidad. Será en torno al mantenimiento homeostático de estas variables que se construya la dinámica cohesiva del SN y su organización interna (sin que, como veremos más adelante, quede totalmente determinada por estas constricciones de adaptabilidad). La dinámica del SN va a estar en última instancia evaluada por su contribución al automantenimiento del organismo, tanto a escala filogenético como ontogenética. La corporización del SN (como sistema jerárquicamente desacoplado) se entiende así en un doble sentido de interfaz interactiva con el mundo y de enraizamiento biológico, mediante el cual el metabolismo sostiene constructivamente al SN y éste contribuye a satisfacer las necesidades interactivas de aquél (intercambio termodinámico con el entorno).

4. Autoorganización y causalidad dinámica en el SN

La autoorganización de los impulsos neuronales en diferentes escalas hace que se generen estructuras intermedias (patrones), y que estos entren a interactuar entre ellos y a organizar la dinámica global del sistema más allá de las interacciones locales de impulsos. Esto nos obliga a distinguir entre niveles dinámicos micro y macro (y entre diversas jerarquías de niveles) y nos permite también hablar de una funcionalidad emergente (e.d. de procesos microscópicos neuronales cuya contribución al automantenimiento del sistema viene marcada por los patrones emergentes que constituyen o en los que participan). Por un lado, la organización de los impulsos neuronales en procesos masivamente paralelos y autoorganizados hace que se diluya la causalidad efectiva de cada uno de los impulsos específicos. Por otro lado, la transformación de patrones globales regenera una causalidad funcionalmente efectiva a un nivel superior. Esto permite hablar de una microdinámica de la que emerge una macrodinámica neuronal funcional. La consecuencia principal de este emergentismo causal es la dificultad de una estrategia localizacionista en el estudio de los procesos cognitivos. El localizacionismo opera haciendo una descomposición estructural del sistema y otra funcional y estableciendo un mapeo entre ambos tipos de descomposición para dar lugar a un explicación causal estructura-función en la que los componentes se agregan linealmente para reconstruir los procesos causales que gobiernan al sistema [8]. Sin embargo, la autoorganización del funcionamiento dinámico que hemos visto no permite llevar a cabo con éxito esta estrategia, ya que la funcionalidad

emerge de la interacción no-lineal entre los componentes [9]. La perspectiva de que la cognición es el resultado de la integración a gran escala de la actividad de conjuntos neuronales distribuidos es cada vez más extendida entre muchos neurocientíficos [10] [11] [12]. La perspectiva más clásica (heredada de la inteligencia artificial tradicional o representacionalista) asume que el cerebro opera en base a módulos funcionalmente específicos de tal manera que la conducta cognitiva es el resulta del procesamiento informacional intramodular y su comunicación inter-modular. Los modelos de integración a gran escala, sin embargo, defienden que la conducta cognitiva es el resultado de patrones globales de oscilación que emergen de la interacción dinámica recíproca entre múltiples conjuntos dispersos de neuronas en el SN. Lo que ha llevado a muchos autores a descartar, prematuramente, el uso del concepto de información semántica (o representación) en el SN.

Un ejemplo paradigmático de autoorganización dinámica a escala neuronal en circuitos locales es el de los CPG. Los CPG (o *central pattern generator*) son conjuntos de neuronas que generan patrones multiestables de impulsos que controlan las neuronas motoras [13] en ausencia incluso de inputs perceptivos (e.d. de forma autogenerada). La generación de patrones estables por el CPG es crucial para constreñir los grados de libertad del sistema muscular y generar un movimiento coherente. Veamos cómo funciona este proceso autoorganizado. Algunas de las neuronas que componen los circuitos de los CPG generan impulsos rítmicos por sí mismas, estas neuronas se acoplan entre sí y con el resto de las neuronas del circuito, generando un patrón global estable. Los CPG son capaces de mantener ese patrón frente a diversas perturbaciones, incluyendo lesiones a partes del circuito. Pero los CPG combinan estabilidad y variabilidad, permitiendo la existencia de diversos patrones estables generados por el mismo circuito. Diversos neuromoduladores y la actividad de neuronas adyacentes al CPG son capaces de provocar cambios de patrones, actuando como selectores de atractores de acuerdo a las necesidades motoras del organismo. Además, a pesar de que los CPG de dos organismos de la misma especie generen patrones similares, sus circuitos suelen ser diferentes. Así pues, diferentes circuitos pueden generar los mismos patrones, y el mismo circuito puede generar diferentes patrones estables; una característica común de los sistemas autoorganizados [14].

5. La aparición de la causalidad informacional en el SN

Con lo que venimos diciendo hasta ahora podemos ya ver que los potenciales de acción (y algunos neuromoduladores) y, sobre todo, los patrones de orden superior que generan al autoorganizarse constituyen un nuevo tipo de observables específicamente neuronales. Cuando se constituye una red formada por varias neuronas en funcionamiento, podemos describir su dinámica a través de esos observables. Mientras consideremos a esta red neuronal como un sistema aislado (por ejemplo, si analizamos su comportamiento en una placa de Petri) tales observables no constituyen sino una forma útil de describir de manera abreviada una mucho más compleja dinámica metabólica subyacente. Pero en el marco

de la actividad vital de un organismo situado en un entorno, estos observables neuronales, como hemos visto, generan un nuevo nivel dinámico de carácter funcional. Como vamos a ver, este nuevo nivel tiene dos características especiales:

1. La característica de todo o nada de los impulsos nerviosos permite una combinabilidad estable de los mismos, que unida a las características de red de la matriz de conectividad y la acción de los neuromoduladores, genera un dominio composicional (los impulsos pueden organizarse secuencialmente en el tiempo), recurrente (la estructura de red permite circularidad) y recursivo (los impulsos operan sobre sí mismos a través de los neuromoduladores que activan[4]. A su vez estos observables pueden autoorganizarse en patrones estables y generar dominios composicionales a niveles superiores.
2. La propiedad fundamental de los observables neuronales es su *potencialidad* para una *causalidad formal* desligada de la causalidad energéticamente determinada sobre la que se propagan esas mismas señales. Del mismo modo que, por ejemplo, los impulsos eléctricos que viajan en los cables de redes informáticas pueden causar cambios en las terminales no en virtud de la energía eléctrica que llevan, sino de la secuencia de cambios de amplitud y frecuencia. La causalidad motora de los impulsos nerviosos no viene determinada por la energía electroquímica que constituye los potenciales de acción, sino por el orden sequencial de impulsos cuya estructura desencadena una serie de reacciones en las células musculares de tal modo que estas movilizan la energía metabólica para producir movimiento. Respecto a la percepción, la peculiaridad de los observables neuronales es que sus efectos causales son definidos por la propia dinámica neuronal que proyecta sobre la percepción una serie de constricciones selectivas de acuerdo a lo que desde un nivel psicológico se describen como estados de atención, anticipación de sentido, etc.

La constitución del SN como un dominio autónomo, jerárquicamente desacoplado, constituido por una dinámica recurrente en constante autoorganización y con capacidad para una causalidad formal, nos acerca irremediablemente al concepto de información. Sin embargo, nos queda por especificar qué es lo que añade el concepto de información a lo que venimos explicando. El factor clave va a ser la asignación de contenido (perceptivo o instructivo).

Para evitar caer en las paradojas homunculares y externalistas de un concepto de información no naturalizado, debemos evitar la tentación de recurrir a las metáforas derivadas de las tecnologías de la computación y la comunicación más allá de lo estrictamente necesario y naturalizable. El punto de unión es, sin duda, la posibilidad de una causalidad formal. Sin embargo la causalidad

[4] Por eso, la diferencia fundamental entre los observables del SN y otro tipo de señales, como las hormonas (moléculas que vehiculizan formas de coordinación intercelular) además de una muchísima mayor rapidez de transmisión, es la capacidad de "procesamiento" recursivo (señales operando sobre señales) y que confiere al SN un carácter "cuasi-sintáctico" en relación a la dinámica de los procesos metabólicos.

informacional en las tecnologías de la comunicación depende siempre de un sistema cognitivo natural que interpreta las señales. Se trata (en nuestro caso, el de la caracterización del concepto de información en el SN) de explicar precisamente el funcionamiento de este último, e.d. de parar la regresión al infinito de sujetos interpretantes (en forma de homunculos cerebrales) que justifique dicha causalidad informacinal, enraizándola en la propia organización dinámica de la actividad neuronal.

La única forma de conseguir que el concepto de información sea naturalizado es (siguiendo a Bickhard [15]) buscando una noción de información (aunque este autor hablará de representación) que permita detección del error por parte del sistema que maneja la información, lo que a su vez exigirá naturalizar la noción de normatividad, de tal modo que ésta sea accesible para el organismo mismo. La normatividad se convierte así en el principio en virtud del cual se establece *la evaluación del error* como proceso *causalmente* efectivo en el mantenimiento de la identidad del sistema. De este modo un concepto naturalizado de información pasa por comprender el origen de los valores regulativos de un organismo. Sólo en el contexto de la interacción biológica sistema-entorno, en el que se generan los valores regulatorios de la agencialidad a través de los procesos autoorganizativos de selección de señales, puede comprenderse el concepto de información en los sistemas cognitivos naturales. Los procesos informacionales en el SN no son meras correlaciones con estados de cosas en el mundo, ni son meras constricciones impuestas sobre los procesos autoorganizativos.

Nos encontramos por tanto con que los factores clave para una noción naturalizada de información en el SN vienen de la mano de cómo se integra la información causalmente en la organización del SN (en relación a su creación y su capacidad de constricción de la dinámica neuronal) y cómo adquieren estos procesos un contenido semántico (referencial) en relación a una normatividad internamente generada por el sistema (y que permita, por tanto, la corrección de error). Veamos cómo podemos reconstruir una noción de información que satisfaga estas condiciones desde la perspectiva dinámica y naturalizada que venimos defendiendo.

6. Hacia una definición naturalizada de información

En primer lugar, podemos diferenciar dos aspectos del contenido informacional. El primero es el instructivo (un aspecto que comparte con la información genética): decimos que el contenido informacional de una estructura o proceso es instructivo si este proceso es capaz de desencadenar-instruir sistemáticamente y en virtud de su "forma" una serie de procesos funcionales (en última instancia, fuera del SN: acciones musculares o secretoras). Es fácil reconocer este aspecto informacional de la dinámica neuronal, por ejemplo, en las señales que hacen cambiar o que desatan un patrón dinámico específico de un CPG y el consiguiente desencadenamiento de una acción motora. De este modo, en el marco conceptual dinámico, *la información son patrones de señales que actúan como parámetros de control de otros procesos autoorganizados.* En este sentido, la dinámica informacional del

SN puede ser entendida *como el conjunto de procesos que gobiernan los cambios de macroestados generados por los procesos autoorganizativos*, y que hacen que el resultado sean acciones evaluables funcionalmente por el organismo.

Sin embargo, hay otro aspecto más en la información neuronal. Se trata de la dimensión perceptiva de la información. Entendemos por ello que los procesos neuronales están organizados en relación a las interacciones con el entorno en virtud de un contenido intencional o semántico de dichos procesos. Es decir, que la organización informacional del SN no es un conjunto de "instrucciones" motoras aisladas de las interacciones que el sistema establece con su entorno, sino que están vinculadas con los ciclos sensomotores y, sobre todo, con la necesidad de adaptar las respuestas motoras a condiciones específicas del entorno.

Tentativamente podemos ofrecer la siguiente definición de información perceptiva en términos dinámicos:

> Un patrón de señales (**P**) se convierte en informacional dentro de un sistema (**S**) acerca de un estado de cosas **E** ssi: **S** selecciona recurrentemente **P**, y **P** es esencial para el automantenimiento de **S** a través de la ocurrencia de **E**.

Donde el sistema **S** es una estructura automantenida de orden superior a **P**. **S** puede ser un patrón autoorganizado de señales, el SN en su conjunto, todo el organismo e incluso la especie como estructura colectiva (por lo que la selección de **P** puede ocurrir a diferentes escalas locales y globales dentro del SN, en la ontogenia del organismo o a escala evolutiva).

Veamos un ejemplo concreto que ilustre esta definición de información: Una señal lleva información acerca de la presencia de comida para un organismo ssi: las acciones que desencadena esa señal en el organismo contribuyen a su automantenimiento *a través de* la presencia de la comida. Esto permite que la información pueda ser falsa, ya que puede darse el caso de que algo que no es comida desencadene esa señal, que a su vez desencadena la consiguiente interacción cuyo resultado *no* contribuye al automantenimiento del sistema (a través de la presencia de la comida). Además puede darse detección de error si el resultado de la interacción es accesible al organismo, y corrección de error si existen mecanismos de modulación de la conducta en base a la detección del error. Este ejemplo ilustra la forma en que podemos hablar de información perceptiva en un organismo sin caer en la falacia del homúnculo, e.d. sin hacer una descripción de correlaciones estimulo-impulso que desplace (indefinidamente) el problema del significado a un sistema que a su vez tenga que *interpretar* el impulso.

La definición que acabamos de presentar permite hablar de procesos neuronales que operan efectivamente como una causalidad específica (la informacional) para el organismo (ya que éste es capaz de evaluar el contenido semántico del proceso informacional) y no desde la perspectiva de un observador externo que establece el vínculo (en forma de correlación causal o de cualquier otro modo) entre señal y contenido informacional.

A partir de aquí, podremos decir que se crea nueva información neuronal cuando se fija una relación funcional que no existía anteriormente. La información neuronal se puede construir pues de dos formas diferentes en virtud de

la relación entre los aspectos instructivos y perceptivos de la información: *a) Información meramente instructiva (secreto-motora)*: se genera a través de procesos autoorganizativos que se realizan a partir de estructuras genéticamente especificadas. Esta modalidad incluye también lo que podríamos llamar "información sensomotríz básica", que aparece en los actos reflejos, cuando (a raíz de una señal perceptiva) se genera una trayectoria sensomotríz genéticamente especificada, (por lo que el aspecto perceptivo y el instructivo son indisociables). *b) Información perceptiva*: es aquella en la que los aspectos perceptivos se pueden desligar de los instructivos. La información perceptiva se genera cuando se fija una relación que no existía anteriormente entre procesos informacionales instructivos y perceptivos. Podemos hablar de dos aspectos organizativamente diferenciados de la información neuronal (el perceptivo y el instructivo) porque a diferencia de las trayectorias sensomotoras innatas, los procesos de aprendizaje permiten combinaciones variables de información instructiva y perceptiva.

7. Aprendizaje y desarrollo: la creación de información en el sistema nervioso

A diferencia de la información genética en la ontogenia del organismo, la información neuronal es constantemente creada en el SN, y es precisamente esa creación de información a través de la interacción del organismo con su entorno la que va a permitir la progresiva "autodeterminación" del SN. La creación de información es necesaria porque no puede estar genéticamente especificada. La creación de información *a través de procesos autoorganizados* es una tarea permanente del SN[5]. Esto se debe a que la información necesaria para especificar circuitos neuronales sobrepasa enormemente la capacidad del código genético. En el caso del ser humano si los circuitos neuronales estuvieran genéticamente preespecificados eso significaría que 10^6 genes (de los cuales sólo un 20-30 % participan en la construcción del SN) tendrían que almacenar la información suficiente para codificar 10^{14} sinápsis que, a su vez, pueden tomar valores cuantitativos de un espectro considerablemente amplio. Esa reducción de variabilidad (del conjunto de parámetros y conexiones posibles a las que resultan funcionales para el organismo) sólo puede darse a través de procesos autoorganizativos que seleccionen y estabilicen configuraciones específicas, e.d. a través de la generación de información durante la ontogenia del organismo.

El concepto de información que venimos defendiendo es indisociable de una dinámica funcional de prueba y error, en la que se generan y destruyen correlaciones funcionales. Aquí es donde la introducción de una dimensión dinámica (la autoorganizativa) resulta, paradójicamente, fundamental, para construir un

[5] A diferencia de la información genética, la información neuronal es generada en tiempo somático. No hay información neuronal genéticamente definida porque la información neuronal se construye en el ámbito del SN. La información genética (via procesos de desarrollo) especifica los mecanismos (entre ellos la arquitectura básica del SN) que van a permitir a su vez la creación de información neuronal (y, naturalmente, la construcción del organismo como sistema metabólico).

concepto naturalizado de información neuronal. Aparte de las dinámicas sensomotoras más simples, que están genéticamente determinadas, toda funcionalidad interactiva nueva surge como resultado de la interacción recurrente y en red (autoorganizada) de todo un conjunto de neuronas, sin que exista un recorrido sensomotor lineal y unidireccional que determine la conducta del organismo. Desde esta perspectiva la estimulación perceptiva aparece como un conjunto indefinido de perturbaciones (trenes de impulsos) en el estado dinámico de la red neuronal. Algunos de estos trenes de impulsos son seleccionados o reclutados por otro patrón interactivo de orden superior. La estimulación perceptiva es, pues, modulada desde dentro: ciclos sensomotores, internamente generados y regulados, estabilizan un patrón sensorial para integrarlo en el ciclo sensomotor. A su vez, nuevos patrones motores (a través de estructuras CPG) son generados para mantener una serie de invariantes internas. Estos patrones se estabilizan (se seleccionan o activan) en forma de atractores, resonancias, etc. cuando contribuyen al automantenimiento del sistema. También puede suceder que, como resultado de ciclos acción-percepción-acción, el patrón global neurodinámico cambie y el organismo adopte otro tipo de interacción con el entorno. La corporización interactiva del SN hace que este proceso de mantenimiento de invariantes se cierre en ciclos sensomotores con el entorno integrándolos funcionalmente para el mantenimiento de los patrones globales. En una escala de complejidad aún mayor encontramos que la propia dinámica de la red genera procesos de aprendizaje. La forma más básica de los procesos de aprendizaje funcional es la del reforzamiento de las conexiones que han dado lugar a una conducta de valor positivo para el organismo. En este mecanismo son fundamentales los procesos de activación de los neuromoduladores que actúan fijando las rutas funcionales de impulsos, generando así un proceso de *selección* interna [16][6].

8. Conclusiones

A partir de una semántica o referencialidad corporizada en las necesidades adaptativas del organismo y guiada por señales de valor evolutivamente fijadas, va generándose toda una red de procesos autoorganizados en la que unos niveles van seleccionando procesos de niveles inferiores en una jerarquía anillada de procesos informacionales. De este modo los grados de libertad de un sistema (el SN) capaz de un inmenso número de transformaciones de estado (a través de una enorme variabilidad recurrente entre diversos tipos de observables) se van organizando informacionalmente a lo largo del desarrollo ontogenético. Los procesos de aprendizaje, sostenidos sobre la normatividad básica de automantenimiento del organismo, van estabilizando una serie de relaciones perceptivas que, a su vez, estabilizan otros procesos de órdenes superiores, creando una red de relaciones perceptivas siempre revisables, ya que su estabilidad depende de la satisfacción de ciertas correlaciones sensomotoras acopladas al entorno. La organización

[6] La forma general de este proceso es que los neuromoduladores extrínsecos (como la dopamina) regulan la estabilidad de aquellos circuitos (o matrices de conectividad) que generan procesos interactivos de valor adaptativo.

informacional cognitiva es, pues, aquella en la que los procesos neuronales se instruyen a través de señales seleccionadas recursivamente mediante su inserción en ciclos sensomotores que dependen de estructuras específicas del entorno (y generando nuevas relaciones perceptivas con el mismo).

Esta perspectiva, además de ser consistente con la investigación empírica, permite naturalizar el concepto de información en el desarrollo científico de las neurociencias cognitivas. De otro modo el uso heredado de este concepto (proveniente de la tradición de IA clásica) se enfrenta a una serie de problemas que hacen desaparecer, precisamente, aquellas propiedades (intencionales y semánticas) que hace del cerebro uno de los sistemas más complejos y fascinantes de nuestro universo.

Referencias

1. Adrian, E. D. *The basis of sensation; the action of the sense organs.* 1st.edition Christophers London. reprint Hafner N.Y. (1928:1964)
2. Dayan, P y Abbott, L. F. *Theoretical Neuroscience.* MIT Press, Cambridge, MA. (2001)
3. Rieke, F., Warland, D., van Steveninck, R. y Bialek, W. *Spikes. Exploring the Neural Code.* MIT Press. Cambridge, Massachusetts. (1997)
4. Ruiz-Mirazo, K y Moreno, A. Searching for the roots of autonomy: The natural and artificial paradigms revisited. *Comunication and Cognition-Artificial Intelligence,* **17**(3–4): 209–228. (2000)
5. Arhem, P. y Liljenstrom, H. On the Coevolution of Cognition and Cosciousness. *Journal of Theoretical Biology*, **187**: 601–612. (1997)
6. Moreno, A. y Lasa, A. From Basic Adaptivity to Early Mind *Evolution and Cognition* **9**(1): 12–30 (2003)
7. Ashby, W. *Design for a Brain. The origin of adaptive behaviour.* Chapman and Hall, 1978 edition. (1952)
8. Bechtel, W. y Richardson, R. *Discovering Complexity. Decomposition and Localization as strategies in scientific research.* Princeton University Press. (1993)
9. Steels, L. Towards a theory of emergent functionality. En Meyer, J. Wilson, R. (eds.) *Simulation of Adaptive Behavior.* MIT Press, Cambridge MA, pp. 451–461. (1991)
10. Varela, F., Lachaux, J.P., Rodriguez, E. and Marinerie, J.: The Brain Web: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience* (2001) **2**: 229–239.
11. Friston, K.J.: The labile brain (I, II and III). *Phil. Trans. R. Soc. Lond. B* (2000) **355**: 215–265.
12. Tononi, G, Edelman, G.M. and Sporns, O.: Complexity and coherency: integrating information in the brain. *Trends in Cognitive Science* (1998) **2** (12): 474–484.
13. Arshavsky, Y., Deliagina, T. y Orlovsky, G. Pattern Generation. *Current Opinion in Neurobiology*, **7**: 781–789. (1997)
14. Kelso, J.S.A. *Dynamic Patterns. The Self-Organization of Brain and Behavior.* MIT Press, Cambridge, MA. (1995)
15. Bickhard, M. H. Information and Representation in Autonomous Agents. *Journal of Cognitive Systems Research*, **1**(2), (2000)
16. Edelman, G.M. *Neural Darwinism: The Theory of Neuronal Group Selection.* Basic Books. (1987)

Modelo de conductancia sináptica para el análisis de la correlación de actividad entre neuronas de integración y disparo

Francisco J. Veredas y Héctor Mesa

Dpto. Lenguajes y Ciencias de la Computación, Universidad de Málaga
fvn@lcc.uma.es,
http://www.lcc.uma.es

Resumen A lo largo de la vía visual de los mamíferos, las conexiones monosinápticas fuertes generan correlaciones de actividad precisas entre las neuronas presinápticas y las postsinápticas. El grado de precisión de la correlación de actividad entre dos neuronas puede inferirse a partir del ancho del pico del correlograma temporal de su actividad. En las conexiones retinogeniculadas pueden encontrarse picos de correlogramas con una anchura que puede ser inferior a $1\,ms$, indicando una gran precisión en la correlación de actividad. A medida que se avanza a lo largo de la vía visual, la correlación de actividad se vuelve más imprecisa y los picos de los correlogramas muestran mayor anchura. Aislar los parámetros fisiológicos que contribuyen a la forma de estos correlogramas en experimentos fisiológicos resulta difícil. Sin embargo, puede lograrse una primera aproximación mediante un enfoque computacional. En este artículo se presenta un modelo de neurona de integración y disparo diseñado y computacionalmente optimizado para el estudio de los factores fisiológicos que influyen en la correlación de actividad entre pares de neuronas. El modelo se destaca por incorporar una modulación de conductancias basada en funciones alfa por partes, de manera que puede analizarse independientemente la contribución del tiempo de subida y de bajada de la conductancia en la correlación de actividad neuronal.

Palabras clave: neurona de integración y disparo, conductancia sináptica, correlación de actividad, optimización computacional, transformada $\mathcal{Z}$

1. Introducción

A lo largo de la vía visual de los mamíferos, las *conexiones monosinápticas fuertes* [14,19] generan actividad correlacionada entre neuronas presinápticas y postsinápticas. La *precisión temporal* o *sincronía* de estos disparos correlacionados (medida mediante análisis de correlación cruzada) es diferente dependiendo de la región analizada, sea ésta el tálamo o la corteza visual. Mientras que las conexiones retino-geniculadas generan correlogramas con picos muy estrechos

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 207-218, 2006.

(con ancho de pico menor que $1\,ms$), los correlogramas generados por conexiones geniculo-corticales y cortico-corticales presentan usualmente correlogramas con picos de varios milisegundos [19,1]. Son varios los factores que podrían explicar esas diferencias de *precisión temporal*, como, por ejemplo, el curso temporal de los potenciales postsinápticos excitadores (EPSP), la eficacia sináptica, la contribución de conexiones polisinápticas, etc.

Se piensa que la precisión en los disparos correlacionados generados por neuronas presinápticas y postsinápticas desempeña un papel primordial en el desarrollo de la vía visual y su remodelación plástica [8], de manera que el curso temporal de la actividad correlacionada de las neuronas constituye un factor fundamental para que los cambios plásticos se lleven a cabo. Por otro lado, es importante identificar los factores que son responsables de las diferencias en la precisión de los disparos correlacionados a lo largo de la vía visual. Resulta difícil aislar experimentalmente la contribución de cada uno de estos factores, por lo que se hace indispensable recurrir al modelado y la simulación. En este artículo se presenta un modelo de neurona de integración y disparo en el que se han integrado los principales parámetros que permiten, mediante simulación, identificar los factores que influyen en la actividad correlacionada. Uno de los aspectos principales que se analiza es la contribución de las constantes de tiempo de la conductancia sináptica, que a su vez influyen en la evolución temporal de los EPSP. Para ello, el modelo de neurona incorpora un modelo de modulación de conductancia con dos constantes de tiempo, que permite simular independientemente el efecto del tiempo de subida y bajada de los EPSP en la correlación de los disparos neuronales.

1.1. Precisión temporal de disparos correlacionados

Las propiedades de la respuesta neuronal en la vía visual de los mamíferos sufren importantes transformaciones desde la retina hasta la corteza visual primaria. Los *campos receptores* son cada vez más complejos a medida que se avanza en la vía visual [5], las frecuencias de disparo se reducen y las *respuestas neuronales* se vuelven más variables [7]. Los estudios de análisis de correlación cruzada sugieren otra transformación adicional: la correlación de disparos entre neuronas presinápticas y postsinápticas se vuelve menos precisa, es decir, la sincronía entre los disparos disminuye. Mientras que las conexiones retino-geniculadas generan correlogramas con picos muy estrechos, con menos de un milisegundo de anchura a mitad de su amplitud máxima [20], las conexiones geniculo-corticales y cortico-corticales generan correlogramas con picos mucho más anchos [2].

Los trabajos realizados hasta la actualidad no dejan claro cuáles son los factores responsables de estas diferencias en la precisión de los disparos correlacionados. Las conexiones retino-geniculadas difieren de las conexiones geniculo-corticales y cortico-corticales en una gran variedad de factores anatómicos y fisiológicos. Mientras que un aferente retiniano establece más de cien sinapsis con una única célula geniculada, un aferente geniculado o cortical establece usualmente menos de diez sinapsis por cada neurona 'diana'. Los EPSP retino-geniculados son también más amplios, tienen tiempos de subida más cortos, una

menor duración y presentan *jitters* de latencia sináptica menores que los EPSP cortico-corticales [13]. Además, mientras que en una única neurona del NGL convergen de uno a tres aferentes, una neurona cortical de la capa IV puede recibir entrada convergente de alrededor de treinta aferentes geniculados [2]. Las neuronas corticales están también fuertemente interconectadas, mientras que las células geniculadas raras veces establecen conexiones excitadoras las unas con las otras [4].

En los estudios previos sobre el sistema motor y somatosensorial no se llega a un acuerdo sobre si el ancho del pico del correlograma monosináptico está determinado por el tiempo de subida de los EPSP o por la duración de los EPSP. Por otro lado, aunque se ha supuesto que otros parámetros como el ruido sináptico y el umbral de disparo tienen una contribución importante, no han sido explorados con detalle (véase [10] para una revisión). En [21] se ha utilizado el modelo de neurona que se presenta en este artículo (sin optimización mediante transformada $\mathcal{Z}$) para analizar sistemáticamente los distintos factores fisiológicos que influyen en la correlación de actividad entre dos neuronas con conexión monosináptica excitadora. Para ello se simularon circuitos que generan correlogramas 'realistas' que se asemejan a los que se pueden medir a tres niveles distintos en el sistema visual: retina-NGL, NGL-corteza y capa IV-capas II-III de la corteza.

2. Modelo de neurona de integración y disparo

Los modelos de integración y disparo son un caso particular de los modelos simplificados de neuronas, que derivan del trabajo pionero de Louis Lapicque [12]. Diferentes versiones de este modelo original han sido propuestas en la literatura [11,6,18,17]. La forma tradicional de un modelo de neurona de integración y disparo es una ecuación diferencial de primer orden (ec. 1) que contiene un dominio de integración subumbral (donde la neurona integra las entradas $I(t)$) y un potencial umbral de disparo (que no aparece explícitamente en las ecuaciones) para la generación de potenciales de acción.

$$C_m \frac{dV_m(t)}{dt} \ = \ I(t) - \frac{[V_m(t) - V_{rest}]}{R_m}, \tag{1}$$

donde C_m es la capacitancia de la membrana neuronal, V_m es el potencial de membrana, R_m es la resistencia de la membrana, V_{rest} es el potencial de reposo y $I(t)$ es la corriente sináptica de entrada a la neurona. En el modelo empleado en este artículo, la corriente $I(t)$ esta modelada por la ec. 2.

$$I(t) \ = \ \sum_j \omega_j g(\hat{t}_j)[E_{syn} - V_m(t)] + noise, \tag{2}$$

donde ω_j representa el peso de conexión (eficacia sináptica) de la neurona j (presináptica) con la neurona actual (postsináptica), $g(s)$ representa la conductancia sináptica, $\hat{t}_j$ es el tiempo relativo al disparo de la neurona presináptica (teniendo en cuenta la existencia de una latencia y de un *jitter* de transmisión sináptica), E_{syn} es el potencial reverso o potencial de equilibrio sináptico y, por

último, *noise* representa el ruido sináptico [que sigue una distribución normal de parámetros (μ_n, σ_n)].

El modelo de neurona ha sido diseñado con el objetivo, entre otros, de poder determinar la influencia del tiempo de subida y bajada de los EPSP de manera independiente. Tradicionalmente, la conductancia sináptica se modela mediante funciones alfa ($\frac{t}{\tau}e^{1-t/\tau}$) o diferencias de exponenciales negativas ($e^{-t/\tau 1} - e^{-t/\tau 2}$). En ambos casos, no existe un parámetro que permita incidir independientemente en su tiempo de subida o de bajada. Por otro lado, en el modelo que se presenta en este artículo, la conductancia sináptica se ha modelado (ec. 3) mediante una función por partes que consta de dos funciones alfa con diferentes constantes de tiempo, τ_r y τ_d, que afectan de manera independiente a su tiempo de subida y bajada (véase figura 1).

$$g(t) = \begin{cases} \frac{t}{\tau_r} e^{1-\frac{t}{\tau_r}} & si\ t \leq \tau_r \\ \frac{\tau_d - \tau_r + t}{\tau_d} e^{1-\frac{\tau_d-\tau_r+t}{\tau_d}} & si\ t > \tau_r \end{cases} \quad (3)$$

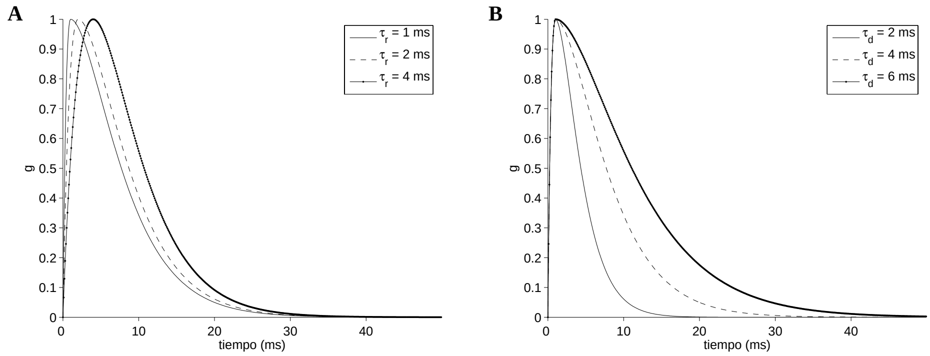


Figura 1. Curvas de conductancia con distintos valores para las constantes de tiempo de subida (A) y de bajada (B).

Para la implementación computacional del modelo de neurona, la ecuación 1 se ha integrado (ec. 4) mediante el método de integración de Euler (*Backward Euler Integration Method*), eligiendo una resolución temporal de $\Delta t = 0,1\,ms$, que asegura la estabilidad del sistema y permite el modelado de fenómenos fisiológicos que, como es el caso del *jitter* de transmisión sináptica, se manifiestan en un orden temporal inferior a $1\,ms$.

$$V_m(t + \Delta t) = \frac{\Delta t}{C_m} I(t) + \left(1 - \frac{\Delta t}{\tau_m}\right) V_m(t) + \frac{\Delta t}{\tau_m} V_{rest} \quad (4)$$

El modelo de neurona de integración y disparo se ha completado incluyendo un período refractario absoluto (del orden de $1,5\,ms$) y un potencial de hiper-

polarización posterior al potencial de acción que se establece como el 20 % del potencial de membrana anterior al disparo de la neurona. Estos factores fisiológicos no se muestran en las ecuaciones anteriores para no mermar su legibilidad.

En el cuadro 1 se muestran los valores habituales de los parámetros del modelo presentes en la literatura sobre modelos de integración y disparo, así como los valores usados en las simulaciones realizadas en el estudio que se presenta en este artículo.

Cuadro 1. Valores estándar de los parámetros habituales en la literatura y valores empleados en las simulaciones de este estudio

Parámetro	**Valor estándar**	**Valor en las simulaciones**
ω_j	$21\,nS$	$6,5\,nS$
C_m	$10\,nF$	$0,1\,nF$
R_m	$10\,M\Omega$	$1\,M\Omega$
V_{rest}	$-65\,mV$	$-70\,mV$
Período refractario absoluto	$2,5\,ms$	$2,5\,ms$
Umbral de disparo	$-40\,mV$	$-40\,mV$
Δt	$0,01\,ms$	$0,1\,ms$
$\tau_m(R_m C_m)$	$7,5-20\,ms$	$1\,ms$
E_{syn}	$0\,mV$	$0\,mV$
τ_r	$0{,}1\,ms$	$0{,}5\,ms$
τ_d	$4\,ms$	$4\,ms$
Ruido sináptico (μ_n, σ_n)	$--$	$(2,6\,nA,\ 0,5\,nA)$
Latencia	$1\,ms$	$1\,ms$
Jitter	$0,5\,ms$	$0,5\,ms$

En la figura 2 se muestra el resultado de una simulación de una conexión monosináptica excitadora que sigue el modelo de neurona presentado en las ecuaciones anteriores, con los parámetros fijados en el cuadro 1: en la parte superior de la figura (A) aparece el correlograma cruzado de los disparos de las dos neuronas (denominadas n_1 y n_2 en la figura), que muestra un pico característico desplazado a la derecha del cero (debido a la latencia sináptica), que indica que n_2 tiende a disparar después de n_1. En la parte inferior de la figura (B) se presenta el potencial de membrana de las dos neuronas simuladas, para una simulación en la que n_2 no recibe ruido sináptico de fondo y su umbral de disparo está desplazado para que la neurona no dispare y se muestren con claridad los EPSP producidos por los disparos de n_1.

3. Influencia de la conductancia en la correlación de actividad

El modelo de neurona de integración y disparo presentado en el apartado anterior se caracteriza por la presencia de una curva de conductancia modelada mediante

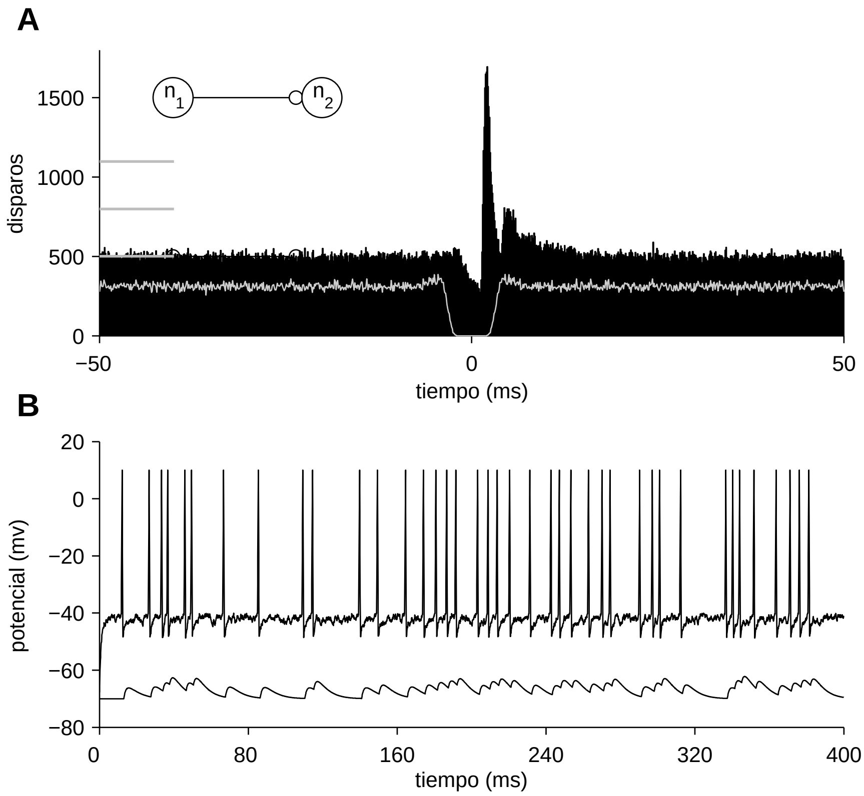


Figura 2. Simulación de una conexión monosináptica excitadora fuerte entre dos neuronas (véase el esquema en la parte superior de la figura). En A se muestra el correlograma cruzado de los disparos (en negro), junto con el autocorrelograma de n_1 (en gris, sólo el perfil); las tres líneas de la izquierda muestran la línea basal de ruido (abajo) y dos niveles usados para medir la anchura del pico (25 % y 50 % de su amplitud máxima) como medida de precisión de correlación de actividad. En B se muestra el potencial de membrana de n_1 (arriba) y n_2 (abajo) para una simulación en la que n_2 está hiperpolarizada y no recibe ruido de fondo.

dos constantes temporales que permiten modificar, independientemente, su tiempo de subida y bajada en las simulaciones. Con objeto de estimar la influencia del tiempo de subida y bajada de la conductancia (y, por tanto, de los EPSP) en la correlación de actividad entre dos neuronas monosinápticamente conectadas, se han lanzado simulaciones donde un sólo parámetro (τ_r o τ_d) varía sistemáticamente en un rango de valores: para cada valor se lanzan 100 simulaciones y se mide el ancho del pico del correlograma obtenido. En la figura 3 se muestran los resultados de estas simulaciones mediante un diagrama de dispersión: sólo la constante de caída de la conductancia (τ_d) va cambiando de valor, entre $0,5\,ms$ y $20\,ms$, afectando a la duración de los EPSP y a la precisión de la correlación de actividad.

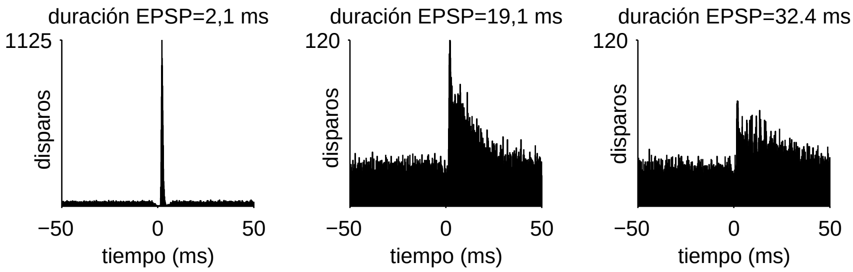
Las simulaciones realizadas para cambios en τ_d, mostradas en la figura 3, arrojan una ratio (máxima anchura / mínima anchura) de 11,87 en la anchura del pico al 25 % de su amplitud máxima (con un coeficiente de correlación lineal de 0,95). Para el tiempo de subida de la conductancia (τ_r) la ratio obtenida es de 5,53 (con coeficiente de correlación lineal de 0,75). Estos datos indican que los tiempos de subida y bajada de la conductancia influyen cada uno de ellos de manera significativa en la precisión de la actividad correlacionada para una conexión monosináptica excitadora y muestran la utilidad del modelo de neurona presentado en este tipo de análisis (véase [21] para completar la revisión de estos resultados).

4. Optimización del modelo de conductancia mediante transformada $\mathcal{Z}$

En el apartado anterior se han presentado resultados de simulaciones de redes constituidas por dos neuronas monosinápticamente conectadas. Para permitir el análisis futuro de redes de mayor tamaño y el estudio de los factores que contribuyen a la correlación de actividad entre pares de neuronas de esas redes, es preciso dar pasos en la optimización computacional del modelo de neurona de integración y disparo. En la mayoría de las simulaciones realistas, incluidas las presentadas en los apartados anteriores, los cambios de la conductancia en cada sinapsis tienen que ser actualizados para cada paso de simulación. Por tanto, la actualización de las conductancias constituye una tarea crítica desde el punto de vista de la eficiencia computacional. El problema básico consiste en computar la conductancia total g_{total} de una manera más eficiente que la mera convolución de los trenes de disparos presinápticos s_i con las funciones de conductancia g_i de todas las sinapsis en el instante t_x:

$$g_{total}(t_x) = \sum_{i=0}^{N} \omega_i \int_0^{t_x} s_i(\tau) g_i(t_x - \tau) d\tau \tag{5}$$

$$s(t) = \sum_{i=0}^{N} \omega_i s_i(t), \tag{6}$$



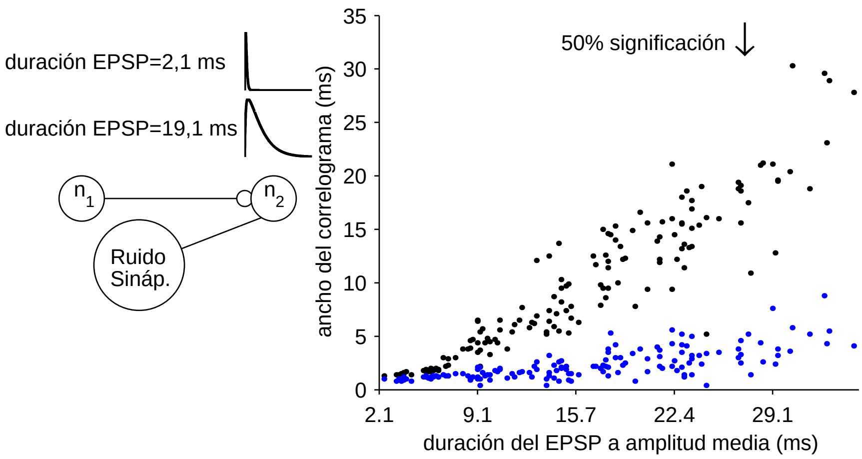


Figura 3. En la parte superior se muestran tres correlogramas (de n_1 a n_2) con diferentes duraciones en los EPSP de n_2; incrementando la duración de los EPSP se consiguen picos monosinápticos más anchos. En la parte inferior de la figura se muestra el diagrama de dispersión, que indica que el pico monosináptico se hace mucho más ancho a medida que aumenta la duración del EPSP (a través de cambios en τ_d); la nube de puntos superior indica ancho de pico al 25 % de su amplitud máxima; en la nube de puntos inferior se muestra en ancho de pico al 50 % de su amplitud máxima; la flecha indica la posición a partir de la cual el 50 % de los correlogramas obtenidos (de los 100 generados para cada valor de τ_d) no tienen picos significativos. A la izquierda del diagrama se muestra un esquema del circuito y ejemplos de los EPSP usados.

donde N es el número total de sinapsis. El problema de la ecuación 5 radica en la necesidad de guardar "la historia"de los disparos presinápticos para poder calcular g_{total} en el instante actual. Normalmente, estos valores se almacenan en *buffers* para que estén disponibles en las siguientes iteraciones de la simulación y cada uno de ellos es empleado en los cálculos en cada paso, lo cual implica la necesidad de realizar numerosas operaciones, viéndose considerablemente afectada la eficiencia computacional. En 1990 Olshausen mostró que la transformada $\mathcal{Z}$ [9,15,3,16] puede ser empleada para acelerar considerablemente los cálculos que implican la ecuación 5.

Partiendo de la ecuación de la conductancia (ec. 3), se pretende, mediante la transformada $\mathcal{Z}$ de la discretización de $g(t)$, llegar a una expresión recursiva que permita expresar $g[n]$ en función de tan sólo unos pocos términos anteriores. Para ello, inicialmente se expresa la conductancia (ec. 3) en una sola ecuación mediante el empleo de la función escalón $[u(t)]$:

$$g(t) = \hat{g}\left([u(t) - u(t-\tau_r)]\frac{t}{\tau_r}e^{1-\frac{t}{\tau_r}} \right.$$
$$\left. + u(t-\tau_r)\frac{\tau_d - \tau_r + t}{\tau_d}e^{1-\frac{\tau_d-\tau_r+t}{\tau_d}}\right) \quad (7)$$

$$u(t) = \begin{cases} 1 & \text{si } t \geq 0 \\ 0 & \text{si } t < 0, \end{cases} \quad (8)$$

$$(9)$$

donde $\hat{g}$ representa la conductancia máxima sináptica. Calculando la transformada $\mathcal{Z}$ de $g[n]$, que es la discretización de $g(t)$ tras el paso $t \rightarrow nT$, se obtiene

$$g[n] = \hat{g}\left(u[n]P_1[n] - u[n-n_r]P_1[n] + u[n-n_r]P_2[n]\right) \quad (10)$$

$$P_1[n] = \frac{n}{n_r}e^{1-\frac{n}{n_r}} \quad (11)$$

$$P_2[n] = \frac{n_d - n_r + n}{n_d}e^{1-\frac{n_d-n_r+n}{n_d}}, \quad (12)$$

siendo $G(z) = \mathcal{Z}\{g[n]\}$, $H(z) = \mathcal{Z}\{g_{total}[n]\}$, $\tau_r \sim n_rT$ y $\tau_d \sim n_dT$.

Usando la propiedad $\mathcal{Z}\{af_1[n] + bf_2[n]\} = aF_1(z) + bF_2(z)$, se calcula la transformada $\mathcal{Z}$ de $g[n]$, por partes, y se obtiene $G(z)$:

$$G(z) = \hat{g}\left[\frac{a_1c_1z}{(z-c_1)^2} - z^{-k_r}\left(\frac{z}{z-c_1} + \frac{a_2c_1z}{(z-c_1)^2}\right)\right.$$
$$\left. + z^{-k_r}\left(\frac{z}{z-c_2} + \frac{a_3c_2z}{(z-c_2)^2}\right)\right], \quad (13)$$

donde $a_1 = \frac{e}{n_r}$, $c_1 = e^{-1/n_r}$, $a_2 = \frac{1}{n_r}$, $a_3 = \frac{1}{n_d}$ y $c_2 = e^{-1/n_d}$.

Por las propiedades de la transformada $\mathcal{Z}$ se tiene que

$$g_{total}(nT) = s(nT) * g(nT) \Leftrightarrow H(z) = S(z)G(z) \tag{14}$$

Aplicando lo anterior a la $G(z)$ obtenida en la ecuación 13, y posteriormente operando y calculando la transformada $\mathcal{Z}$ inversa, se obtiene:

$$\begin{aligned} g_{total}[n] =& p_1 s[n-1] + p_2 s[n-2] + p_3 s[n-3] \\ &+ p_4 s[n-n_r-1] + p_5 s[n-n_r-2] + p_6 s[n-n_r-3] \\ &- r_3 g_{total}[n-1] - r_2 g_{total}[n-2] - r_1 g_{total}[n-3] - r_0 g_{total}[n-4], \end{aligned} \tag{15}$$

siendo

$$p_1 = \hat{g} a_1 c_1 \tag{16}$$
$$p_2 = -2\hat{g} a_1 c_1 c_2 \tag{17}$$
$$p_3 = \hat{g} a_1 c_1 c_2^2 \tag{18}$$
$$p_4 = \hat{g}\left[c_2(1+a_3) - c_1(1+a_2)\right] \tag{19}$$
$$p_5 = \hat{g}\left[c_1^2 - c_2^2 + 2c_1c_2(a_2-a_3)\right] \tag{20}$$
$$p_6 = \hat{g}\left[c_1c_2^2(1-a_2) - c_1^2c_2(1-a_3)\right] \tag{21}$$
$$r_0 = c_1^2 c_2^2 \tag{22}$$
$$r_1 = -2\left(c_1^2c_2 + c_1c_2^2\right) \tag{23}$$
$$r_2 = c_1^2 + c_2^2 + 4c_1c_2 \tag{24}$$
$$r_3 = -2(c_1+c_2) \tag{25}$$

La ecuación 15 es una función recursiva que permite expresar cada valor de g_{total} en el paso n en función de un número reducido de valores anteriores de g_{total} y de los trenes de disparo presinápticos.

Para estimar la ganancia de eficiencia computacional derivada del empleo de la ecuación 15 para la conductancia sináptica, se lanzaron simulaciones de redes completamente conectadas (sin auto-conexiones) de hasta 10 neuronas –con una frecuencia de disparo $\approx 400\,Hz$, máxima para un período refractario de $2,5\,ms$– en un ordenador AMD Athlon^{TM}$(1,4\,GHz)$. El modelo de neurona se implementó en Matlab^{TM}. Cada simulación se repitió tres veces para cada combinación de parámetros (número de neuronas y pasos de simulación) y se obtuvo el promedio de esas tres simulaciones, tanto para la implementación con transformada $\mathcal{Z}$ (ec. 15) como para la implementación basada en la convolución y el uso de *buffers* (ec. 5). En la figura 4 se muestra que la implementación basada en la transformada $\mathcal{Z}$ mejora considerablemente los tiempos de cómputo a partir de un tamaño de red relativamente reducido de ocho neuronas. Así, para 10 neuronas y 6.000 pasos de simulación ($600\,ms$) el tiempo promedio de las tres simulaciones con transformada $\mathcal{Z}$ es un 20 % menor que el de las tres simulaciones con convolución.

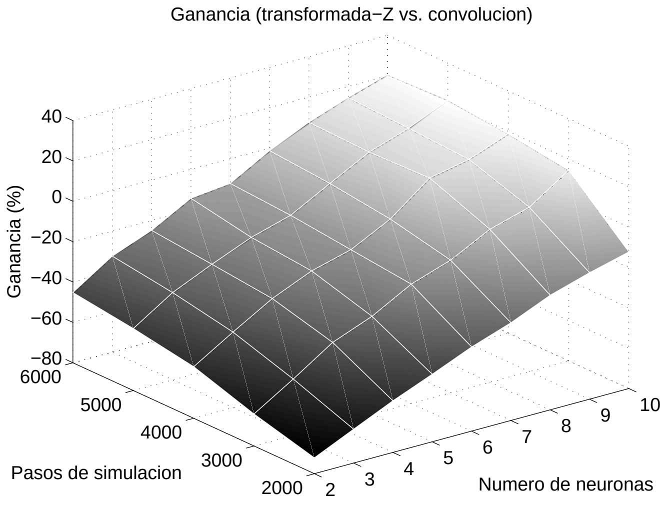


Figura 4. Ganancia de eficiencia de cómputo para simulaciones de una red de hasta 10 neuronas, con conectividad total (sin auto-conexiones) empleando dos modelos de conductancia: mediante convolución de la conductancia con los trenes de disparo (ec. 5) y mediante una ecuación recursiva obtenida mediante la transformada $\mathcal{Z}$ (ec. 15).

5. Conclusiones

El nuevo modelo de neurona de integración y disparo presentado en este artículo permite realizar simulaciones para el análisis de la influencia de diversos factores fisiológicos en la correlación de actividad entre pares de neuronas. La modulación de la conductancia sináptica mediante dos funciones alfa con constantes de tiempo independientes, permite estudiar cómo afecta el tiempo de subida y de bajada de la conductancia, a través de la evolución de los EPSP, en la precisión temporal de los disparos correlacionados en una conexión monosináptica. De las simulaciones de este nuevo modelo de neurona se concluye que el tiempo de subida y bajada de la conductancia influye significativamente sobre la correlación de actividad neuronal en este tipo de conexiones.

La optimización del modelo mediante la transformada $\mathcal{Z}$, permite expresar la ecuación de la conductancia en tiempo discreto como una función recursiva, consiguiéndose un aumento de la eficiencia computacional del modelo (respecto a la obtenida en simulaciones donde la conductancia total se expresa como la convolución de las funciones de conductancia con los trenes de disparo presinápticos) en simulaciones de redes neuronales a partir de un número reducido de neuronas. Esta mejora permitirá reducir considerablemente los tiempos de cómputo para el análisis futuro, mediante simulación, de los factores fisiológicos que contribuyen a la correlación de actividad neuronal en redes de gran tamaño.

Referencias

1. J. Alonso and L. Martínez. Functional connectivity between simple cells and complex cells in cat striate cortex. *Nat. Neurosci.*, 1:95–403, 1998.
2. J. Alonso, W. Usrey, and R. Reid. Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *J. Neurosci.*, 21:4002–4015, 2001.
3. G. Doetsch. *Anleitung zum praktischenGebrauch der Laplace-Transformation und der Z-Transformation.* R. Oldenbourg Verlag, Munich, 1967.
4. M. Friedlander, C. Lin, and S. Sherman. Structure of physiologically identified X and Y cells in the cat's lateral geniculate nucleus. *Science*, 204:1114–1117, 1979.
5. D. Hubel and T. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond)*, 160:106–154, 1962.
6. J. Jack, D.Ñobel, and R. Tsien. *Electric current flow in excitable cells.* Oxford University Press, Oxford, United Kingdom, revised paperback (1983) edition, 1975.
7. P. Kara, P. Reinagel, and R. Reid. Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron*, 27:635–646, 2000.
8. L. Katz and C. Shatz. Synaptic activity and the construction of cortical circuits. *Science*, 274:1133–1138, 1996.
9. J. Köhn and F. Wörgötter. Employing the Z-Transform to Optimize the Calculation of the Synaptic Conductance of NMDA and Other Synaptic Channels in Network Simulations. *Neural Computation*, 10:1639–1651, 1998.
10. P. Kirkwood. On the use and interpretation of cross-correlations measurements in the mammalian central nervous system. *J. Neurosci. Methods*, 1(2):107–132, 1979.
11. B. Knight. Dynamics of encoding in a population of neurons. *J. Gen. Physiol.*, 59:734–766, 1972.
12. L. Lapicque. Recherches quantitatifs sur l'excitation electrique des nerfs traitée comme une polarisation. *J. Physiol. Paris*, 9:620–635, 1907.
13. H. Markram, J. Lubke, M. Frotscher, A. Roth, and B. Sakmann. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol.*, 500:409–440, 1997.
14. D. Mastronarde. Two clases of single-input X-cells in cat lateral geniculate nucleus. II. Retinal inputs and the generation of receptive-field properties. *J. Neurophysiol.*, 57:7757–7767, 1987.
15. B. Olshausen. Discrete-time difference equations for simulating convolutions. Technical report, California Institute of Technology, Pasadena, 1990.
16. A. Oppenheim and R. Schafer. *Digital-signal processing.* Prentice Hall International, London, 1975.
17. R. Stein. The frequency of nerve action potentials generated by applied currents. *Proc. Roy. Soc. Lond. B*, 167:64–86, 1967.
18. R. Stein. Some models of neuronal variability. *Biophys. J.*, 7:37–68, 1967.
19. W. Usrey, J. Reppas, and R. Reid. Paired-spike interactions and synaptic efficacy of retinal inputs to the thalamus. *Nature*, 395:384–387, 1998.
20. W. Usrey, J. Reppas, and R. Reid. Specificity and strength of retinogeniculate connections. *J. Neurophysiol.*, 82:3527–3540, 1999.
21. F. Veredas, F. Vico, and J. Alonso. Factors determining the precision of the correlated firing generated by a monosynaptic connection in the cat visual pathway. *J. Physiol.*, 567(3):1057–1078, 2005.

RNA + SIG:
Sistema automático de valoración de viviendas

Noelia García Rubio, Matías Gámez Martínez y Esteban Alfaro Cortés

Facultad de Ciencias Económicas y Empresariales, Universidad de Castilla-La Mancha,
Plaza de la Universidad, 1, 02071 Albacete, España
{Noelia.Garcia, Matias.Gamez, Esteban.Alfaro}@uclm.es

Resumen. El objetivo de este trabajo es construir un sistema automático de valoración de viviendas. Para ello se han integrado modelos de Redes Neuronales Artificiales con un Sistema de Información Geográfica; herramientas que han demostrado su potencial en el ámbito económico. Los modelos de Redes Neuronales utilizados en este trabajo son el Perceptrón Multicapa, las Redes de Función de Base Radial y los Mapas Autoorganizados o Mapas de Kohonen. Los dos primeros suponen una alternativa atractiva a otras técnicas estadísticas más tradicionales en materia de regresión para la estimación del precio de la vivienda. Por otra parte, los Mapas de Kohonen, junto con el Perceptrón Multicapa se han utilizado como herramienta de clasificación en tareas intermedias como la determinación de la calidad de las viviendas.

Palabras clave: Redes Neuronales Artificiales, Sistemas de Información Geográfica, Precio de Vivienda

1 Introducción

El objetivo de este estudio es mostrar cómo diferentes modelos de Redes Neuronales Artificiales (RNA) se pueden combinar con un Sistema de Información Geográfica (SIG) para constituir una herramienta potente en la investigación en Economía; concretamente en el diseño de sistemas de valoración automática de viviendas, así como en otras tareas complejas relacionadas con el mercado inmobiliario como es la determinación objetiva de la calidad de una vivienda, testigo fundamental a tener en cuenta en cualquier tasación.

Los modelos de RNA utilizados son el Perceptrón Multicapa (MLP), las Redes de Función de Base Radial (RBF) y los Mapas Autoorganizados o Mapas de Kohonen (SOFM). Los dos primeros suponen una alternativa atractiva a otras técnicas estadísticas tradicionales en materia de regresión y clasificación supervisada, mientras que los SOFM están especialmente diseñados para tareas de agrupamiento.

El trabajo se estructura de la siguiente forma. En primer lugar, se describirá el problema a resolver, estimación de precios de viviendas libres en la ciudad de Albacete, detallando la información utilizada en la aplicación práctica, así como el tratamiento previo de la muestra utilizada. Tras ello, se pondrán en marcha los modelos de RNA diseñados para la tarea, resaltando los resultados más interesantes. Posteriormente se

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 219-230, 2006.

mostrará el programa diseñado en el entorno gráfico SciViews de R para combinar los resultados de la mejor red neuronal junto con el sistema de información geográfica dando lugar a un verdadero sistema de valoración automática de las viviendas de la ciudad de Albacete. Por último, se expondrán las principales conclusiones derivadas del estudio.

2 Descripción del problema

Como se ha puesto de manifiesto, el objetivo de esta investigación es la obtención de un sistema automático de valoración que, a partir de la localización de una vivienda y de algunos otros testigos, proporcione de manera objetiva una estimación de su precio de mercado. El punto de partida es la obtención de la muestra que, dadas las carencias de las fuentes de información estadística oficiales, necesariamente nos hace recurrir a las agencias inmobiliarias. La muestra obtenida, tras un proceso largo y tedioso, consta de 591 registros, correspondientes a otras tantas operaciones de compra-venta realizadas en 2002 en Albacete, con información sobre las siguientes variables:

- Tipo de vivienda (TIPO), variable categórica con dos valores, 0 en el caso de un piso y 1 en el caso de vivienda unifamiliar.
- Localización, calle y número de la finca en la que se encuentra la vivienda. Esta información se transforma en las variables COORDX y COORDY mediante la localización del punto exacto en el mapa georreferenciado[1] de la ciudad.
- Antigüedad (ANTI), expresada en años.
- Superficie (SUP), expresada en metros cuadrados útiles.
- Dormitorios (DORM), número de habitaciones aparte del salón.
- Baños (BAÑO), variable numérica resultado de sumar un punto por baño completo y medio por cuarto de aseo.
- Ascensor (ASCEN), variable categórica con valor 0 en caso de no contar con ascensor y 1 en caso afirmativo.
- Terraza (TERRA), variable categórica con valor 1 en caso de contar con terraza de superficie igual o mayor de 15 m^2 y 0 en caso contrario.
- Calefacción (CALE), variable categórica con valor 0 en caso de no contar con sistema de calefacción y 1 en caso afirmativo.
- Calidad (CALID), variable categórica con tres clases:
  - Mala, para viviendas antiguas construidas con malos materiales y en malas condiciones de conservación y habitabilidad.
  - Estándar, para viviendas seminuevas construidas de acuerdo a niveles estándar de calidad o para viviendas antiguas rehabilitadas.
  - Muy buena, para viviendas nuevas de primera calidad o antiguas rehabilitadas con elementos por encima de lo que puede ser considerado estándar.

[1] Este mapa fue facilitado por el Instituto de Desarrollo Regional, concretamente, por la sección de Teledetección y SIG, después de contar con la autorización por parte del alcalde de la ciudad de Albacete para su uso y aprovechamiento en esta investigación.

- Garaje (GARAJ), variable numérica que cuenta el número de plazas de garaje.
- Trastero (TRAST), variable categórica con valor 1 en caso de contar con trastero y 0 en caso contrario.
- Distancia al centro (GABLOD), como información adicional a la suministrada por las agencias, el uso del SIG ha permitido incluir una variable considerada importante como es la distancia al centro de la ciudad[2].
- Precio total de la vivienda [3](PRET). A partir de esta información se construye una nueva variable, PREM2, como cociente entre el precio total de la vivienda y su superficie útil. Estas dos son las variables dependientes que constituyen la salida de los diferentes modelos de red que se van a diseñar.

A continuación se reproduce el plano electrónico de la ciudad, donde se recogen las 591observaciones muestrales.

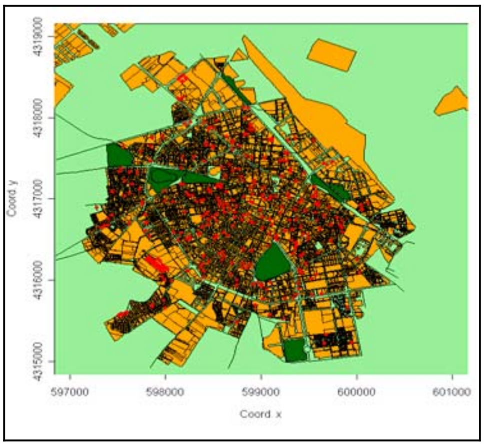


Fig. 1. Plano de la ciudad de Albacete

Hay que señalar que, en principio, se pretendía utilizar más información de entrada, como es la orientación de la vivienda o la presencia de zonas comunes de recreo como piscina, jardines, etc. Sin embargo, en muchos casos esa información fue omitida por las agencias, por lo que finalmente tuvo que ser desechada. En el resto de variables que presentan valores omitidos en un porcentaje razonable se ha optado por realizar el completado de las mismas mediante la técnica de los *k-vecinos más próximos* en el caso de variables cuantitativas y mediante RNA[4] (MLP y SOFM) para tareas de clasificación en el caso de variables cualitativas.

Mostramos, a continuación, los resultados en el completado de la variable calidad.

[2] Como centro de la ciudad se ha tomado la Plaza de Gabriel Lodares, después de descartar otros puntos como la Plaza del Altozano o el Ayuntamiento. La razón es que estos “centros históricos” se han quedado desplazados ante el crecimiento de la ciudad que en dirección nordeste ha quedado limitado por las vías del tren.

[3] Merece la pena destacar que no se trata de valores de oferta sino que son precios de mercado resultado de operaciones de compra-venta llevadas a fin.

[4] Las estimaciones de los modelos de RNA se realizan mediante el programa TRAJAN.

Esta variable recoge una gran cantidad de información y, en muchas ocasiones, ésta es además muy subjetiva. En veintitrés casos muestrales la agencia no había asignado ningún nivel de calidad y además tampoco había información suficiente (reformas, estado de suelos, carpintería, ventanas,…) para asignarlo nosotros mismos. Los 562 casos con información sobre calidad se han utilizado para desarrollar un modelo de RNA capaz de estimar satisfactoriamente los casos perdidos a partir del resto de testigos. Para poder elegir entre los diversos modelos de tipo MLP y SOFM probados y, posteriormente, validar los modelos elegidos, la muestra disponible se ha dividido en tres grupos: entrenamiento (288 casos), validación (137 casos) y test (137 casos). Tras numerosas pruebas con redes de tipo MLP y SOFM, se han seleccionado los dos mejores modelos en cada uno de los tipos: una red MLP 14:14-6-3-1 (catorce neuronas en la capa de entrada, una capa oculta con seis nodos y una neurona de salida para cada nivel de calidad) y una red SOFM 14-49[5] (catorce neuronas en la capa de entrada y cuarenta y nueve en la capa de competición). La tabla 1 muestra los resultados de las dos redes seleccionadas.

Tabla 1. Matriz de confusión para los modelos MLP y SOFM seleccionados

			CLASE ASIGNADA						
			MLP			SOFM			
			Mala	Estándar	Muy buena	Mala	Estándar	Muy buena	Sin clase
CLASE REAL	ENTREN.	Mala	43	1	0	30	10	0	4
		Estándar	7	203	2	5	198	5	4
		Muy buena	0	8	24	2	15	14	1
	VALID.	Mala	14	5	0	13	6	0	0
		Estándar	5	96	0	6	89	5	1
		Muy buena	2	7	8	0	13	4	0
	TEST	Mala	17	3	0	14	5	1	0
		Estándar	5	91	3	1	90	3	5
		Muy buena	0	3	15	0	9	8	1

De la tabla 1 merece la pena destacar la forma diagonal de las matrices de confusión, indicativo de que apenas hay confusión entre clases extremas. Centrándonos en

[5] Aunque, en principio, los mapas auto-organizados están diseñados para realizar tareas de agrupamiento no supervisado, también es posible utilizarlos para tareas de clasificación en clases conocidas a priori. Una vez entrenada la red es posible etiquetar cada una de las neuronas de la capa de competición. Se puede, por tanto, calcular los porcentajes de casos bien clasificados como se acostumbra a hacer en clasificación supervisada. En este trabajo se ha impuesto como condición para etiquetar una neurona que al menos un 50% de los casos para los que una misma neurona resulta ganadora deben pertenecer a la misma clase. Al imponer esta restricción se trata de mantener un "tira y afloja" entre el número de neuronas que quedan sin etiquetar y la confianza que inspiran las neuronas etiquetadas. Como se puede ver en la tabla hay algunos casos sin nivel de calidad asignado, esto es porque han "caído" en alguna neurona sin etiquetar.

el conjunto de test para comprobar la capacidad de generalización de la redes, el MLP presenta un porcentaje de casos correctamente clasificados de casi el 90%, mientras que el mapa de Kohonen sólo ha alcanzado el 82%. Por tanto, la tarea de completado se ha realizado finalmente con las predicciones proporcionadas por el modelo MLP[6].

3 Estimación del precio de vivienda libre

Una vez descrita la muestra, entramos en el objetivo principal de la investigación, estimación del precio de viviendas libres en la ciudad de Albacete. Esta tarea se ha abordado desde dos puntos de vista, el precio por metro cuadrado y el precio total y para cada uno de ellos se han puesto en marcha un buen número de redes tipo MLP y RBF. La tabla 2 muestra los resultados de cada uno de los modelos seleccionados.

Tabla 2. Estadísticos de regresión

	PRECIO METRO CUADRADO		PRECIO TOTAL	
	MLP	RBF	MLP	RBF
Media muestral	1285,9390	1292,1970	139967,7	135204,4
Desviación típica muestral	296,0914	301,5658	46032,14	47511,7
Media error	18,3310	-24,8083	-271,8061	-1183,905
Desviación típica error	139,3686	153,6088	13049,79	17601,32
Media error valor absoluto	98,6936	116,802	9085,213	13220,65
Error medio relativo	7,6748	9,0390	6,4909	9,7783
D. T. Ratio	0,4707	0,5093	0,2835	0,3705
Correlación (obj. y est.)	0,8833	0,8606	0,9591	0,9317
R^2	0,8129	0,7406	0,9196	0,8628

El estadístico de regresión más significativo es la desviación típica del error. Si el valor del estadístico no es mejor que la desviación típica muestral, entonces la red no predice mejor que la simple media. El D. T. ratio compara en forma de cociente la desviación típica del error de predicción y la muestral, de forma que un valor significativamente por debajo de 1 será indicativo de buen ajuste.

Examinando los estadísticos anteriores, se concluye que los mejores resultados se obtienen para el MLP estimando el precio total de las viviendas. Por ello, analizamos más detalladamente el proceso de diseño y entrenamiento de este modelo.

La arquitectura del MLP propuesto se ha seleccionado tras un gran número de pruebas. El número de neuronas en las capas de entrada y salida están determinadas por el número de variables independientes y dependientes respectivamente. Por otra parte, el número de unidades en la capa oculta se ha elegido tratando de construir la

[6] Es importante señalar que en una comparación caso por caso, el grado de acuerdo entre los dos modelos seleccionados es del 75%.

red más sencilla posible. Finalmente, la estructura de la red queda como 14:16-5-1:1, es decir, una capa de entrada con catorce neuronas, preprocesadas en dieciséis unidades[7], una capa oculta con cinco neuronas y una de salida con una única neurona.

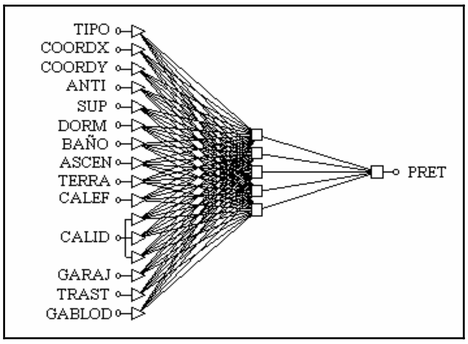


Fig. 2. Arquitectura de un MLP 14:16-5-1:1

A continuación detallamos el proceso de entrenamiento de la red. Éste comienza con la división de la muestra en tres subconjuntos: entrenamiento (50%), validación (25%) y test (25%). Las funciones de activación seleccionadas son la lineal para la capa de entrada y la sigmoidea para las unidades de la capa oculta y de salida. Tras la inicialización aleatoria de la matriz de pesos se lleva a cabo el entrenamiento mediante la regla Delta-bar-Delta[8], con la suma de cuadrados de los errores como función de error. Los detalles del algoritmo se han fijado como sigue:

- Número máximo de iteraciones: 2500
- Tasa inicial de aprendizaje: 0,001
- Incremento: 0,07
- Decaimiento[9]: 0,5

Una vez completado el proceso de entrenamiento, tras 2346 iteraciones, se procede

---

[7] Todas las variables se han preprocesado antes de ser introducidas en la red. Las variables cuantitativas se han normalizado para tomar valores en [0-1] y las cualitativas han sido codificadas en una neurona con dos estados, salvo el testigo CALIDAD, que se ha codificado según el método 1-de-N y, por tanto, necesita tres neuronas.

[8] La regla de aprendizaje Delta-Bar-Delta, propuesta en [6], constituye una modificación del algoritmo Backpropagation estándar. El objetivo es la aceleración de la convergencia del proceso de aprendizaje teniendo, como idea base, la observación de que la superficie de error puede tener gradiente diferente a lo largo de la dirección de cada peso. En esta situación puede ser deseable considerar tasas de aprendizaje distintas para cada parámetro ajustable de la red y, además, que estas tasas sean adaptativas en el tiempo. Así, la tasa de aprendizaje debería ser más elevada cuando los cambios de pesos en pasos consecutivos se producen en la misma dirección y debería disminuir cuando los cambios presenten signo opuesto.

[9] Hay que señalar que el incremento de la tasa de aprendizaje se produce de manera lineal, mientras que el decaimiento es exponencial. Esto es así para evitar que la tasa crezca demasiado rápido y, sin embargo, permitir que cuando sea necesario la tasa pueda ser reducida rápidamente y además garantizar su signo positivo.

a la validación del modelo. Volviendo a los resultados de la tabla 2, en el subconjunto de test, el porcentaje de varianza explicada está muy próximo al 92%, con un coeficiente de correlación entre output de la red y el objetivo de 0,96 y una media del error en valor absoluto de 9085€.

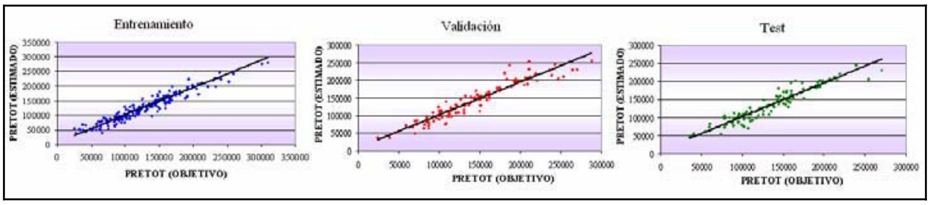


Fig. 3. Correlación entre precio objetivo y precio estimado

Sin embargo, conviene puntualizar algunos aspectos. El primero de ellos es que si bien un elevado coeficiente de correlación no implica coincidencia entre los precios objetivo y los proporcionados por la red, sí podemos hablar de casi coincidencia a la vista de los gráficos en la figura 3. En estos gráficos se puede comprobar que las líneas de tendencia coinciden prácticamente con las diagonales de los cuadros, o en otras palabras, los puntos se encuentran muy cercanos a la línea donde coinciden precios objetivo y estimados. Además, como se puede observar en la figura 3, hay muy pocos puntos lejos de la línea de tendencia y coinciden con unos precios extremadamente altos. En este sentido, resulta interesante analizar el histograma de los errores en la figura 4 para concluir que después de eliminar tan solo seis valores extremos, los resultados cambian significativamente. Para ser más exactos, la media del error en valor absoluto se reduce hasta los 7811€, lo que supone un error medio relativo del 5,65%.

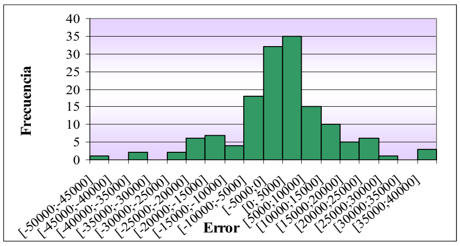


Fig. 4. Histograma de errores en conjunto de test

Validado el modelo y considerando satisfactoria la actuación de la red, podemos cuantificar la contribución de cada variable a dicha actuación mediante un análisis de sensibilidad. La idea es analizar el comportamiento de la red como si un determinado input no estuviera disponible.

La tabla 3 muestra, para cada variable, el cociente entre el error de la red como si esa variable no estuviera disponible y el error con todas las variables de entrada disponibles. Por tanto, un ratio de uno o menor implica que la eliminación de esa variable no tiene efecto negativo sobre el comportamiento de la red.

Tabla 3. Análisis de sensibilidad

Ranking	Variable	Ratio	Ranking	Variable	Ratio
1	Superficie	2,6978	8	Calidad	1,4074
2	Calefacción	2,1082	9	Terraza	1,2087
3	Ascensor	1,9416	10	Antigüedad	1,1859
4	GabLod	1,7586	11	Coord Y	1,1514
5	Tipo	1,6852	12	Dormitorios	1,1129
6	Garaje	1,5713	13	Coord X	1,0972
7	Trastero	1,4987	14	Baño	1,0486

A la vista de los resultados anteriores puede concluirse que todas las variables independientes pasan la prueba de sensibilidad.

Por último, en cuanto a los resultados de la red, analizamos algunas de las superficies de respuesta más interesantes. En la figura 5 se muestra la respuesta de la red ante cambios en la edad y la distancia al centro conjuntamente, así como ante la superficie de la vivienda y su edad.

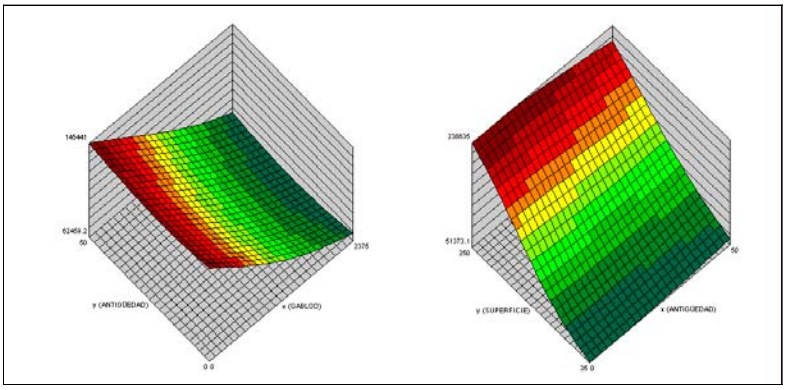


Fig. 5. Superficies de respuesta de la red ante los testigos antigüedad y distancia al centro y superficie y antigüedad.

Como era de esperar, la respuesta de la red ante la distancia al centro (GABLOD) es perfectamente inversa. Sin embargo, es interesante analizar la respuesta de la red ante cambios en esa variable conjuntamente con la edad de la propiedad. Así, se puede ver que en las zonas más cercanas al centro, los precios más elevados parecen corresponder a las viviendas más antiguas, mientras que alejándonos del centro se observa un comportamiento no lineal con precios más elevados en los extremos del testigo antigüedad. En cuanto a la superficie, la respuesta es, tratándose de precio total, lógicamente positiva. Pero además, se puede observar que según sea el tamaño de la vivienda cambiará el patrón de respuesta de la red ante el testigo antigüedad. Así, en las viviendas más pequeñas, el precio tiende a aumentar con la edad. Por otra parte, si se consideran las viviendas de superficie media, el precio decrece al aumentar la edad pero sólo hasta cierto nivel a partir del cual vuelve a incrementarse. Por último, en las viviendas de mayor tamaño, el precio responde negativamente ante

incrementos en la edad.

A la vista de los resultados anteriores, parece conveniente detenerse un poco más en el análisis de la influencia del testigo antigüedad sobre el precio (Fig. 6). En ella se puede ver que, manteniendo el resto de testigos constantes en valores medios, el mínimo valor del precio total estimado corresponde a una antigüedad de 20 años. Este resultado es conforme a los establecidos en [2] acerca de la reversión que tiene lugar en la relación entre antigüedad y precio de las viviendas, dado que encuentran que la relación negativa persiste únicamente hasta los quince o veinte años. La razón de este comportamiento fue teorizada en [11], argumentando que a partir de cierta edad, el precio de la vivienda comienza a aumentar debido al incremento del valor del suelo sobre el que está construida la vivienda. No podemos dejar de señalar en este punto, la conveniencia de utilizar modelos flexibles como las RNA, que permitan recoger las relaciones no lineales, como la anterior, entre testigos de la vivienda.

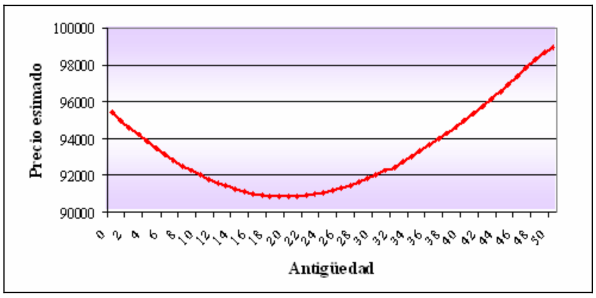


Fig. 6. Respuesta de la red ante el testigo antigüedad.

4 Integración del SIG y el modelo de RNA para la estimación del precio de la vivienda

Estimado y validado el modelo de predicción de precios de vivienda puede ser utilizado para la valoración de viviendas fuera de la muestra. Para obtener un sistema más operativo se han integrado la red estimada y el sistema de información geográfica de Albacete en el entorno gráfico de R SciViews. Además se ha personalizado la aplicación mediante la creación de los botones necesarios para que cualquier usuario pueda visualizar cualquier punto del plano de Albacete y estimar precios en él sin más que introducir los valores de los distintos testigos de la vivienda. La figura 7 muestra la página de inicio de la aplicación con los botones mencionados.

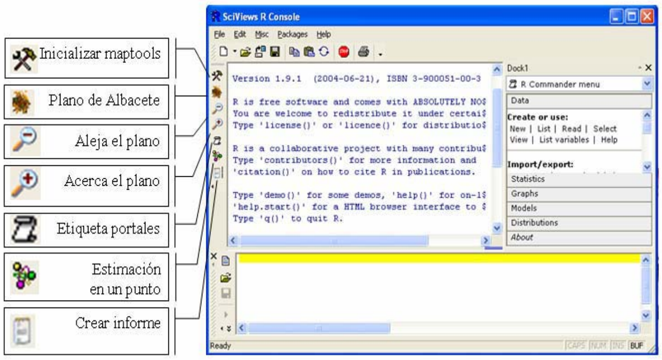
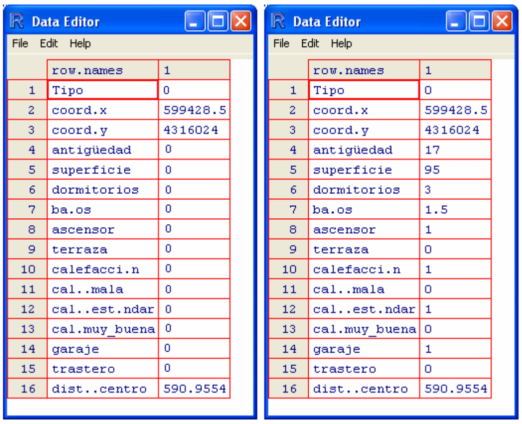


Fig. 7. Página de inicio de la aplicación y descripción de los botones principales

El primer paso consiste en inicializar “maptools”, con ello se llama a las diferentes capas del sistema de información geográfica (calles, manzanas, portales y parcelas) y las herramientas para su manipulación. Hecho esto, cargamos el plano haciendo clic en el botón correspondiente y sobre el mapa nos podremos mover alejándonos y acercándonos para localizar la zona donde se encuentra la vivienda que se pretende valorar. Localizada la vivienda, hacemos clic en el botón “Estimación en un punto” y con él pinchamos en el punto del mapa correspondiente. Aparece entonces un cuadro de edición de datos (Fig. 8, izquierda) en el que se deben introducir las características de la vivienda. Nótese que, gracias al SIG, la localización y la distancia al centro aparecen automáticamente.



	row.names	1
1	Tipo	0
2	coord.x	599428.5
3	coord.y	4316024
4	antigüedad	0
5	superficie	0
6	dormitorios	0
7	ba.os	0
8	ascensor	0
9	terraza	0
10	calefacci.n	0
11	cal..mala	0
12	cal..est.ndar	0
13	cal.muy_buena	0
14	garaje	0
15	trastero	0
16	dist..centro	590.9554

	row.names	1
1	Tipo	0
2	coord.x	599428.5
3	coord.y	4316024
4	antigüedad	17
5	superficie	95
6	dormitorios	3
7	ba.os	1.5
8	ascensor	1
9	terraza	0
10	calefacci.n	1
11	cal..mala	0
12	cal..est.ndar	1
13	cal.muy_buena	0
14	garaje	1
15	trastero	0
16	dist..centro	590.9554

Fig. 8. Editor de datos para la introducción de las características de la vivienda

Introducimos como ejemplo los datos de una vivienda que hemos localizado en la calle Cristóbal Lozano, 10. Suponemos que se trata de un piso de 17 años de antigüedad, con una superficie de 95m2, 3 dormitorios, un cuarto de baño completo y un aseo, con ascensor, sin terraza, con calefacción, calidad estándar, una plaza de garaje y sin trastero. Como resultado, el precio estimado aparece en un recuadro sobre el punto del mapa donde se encuentra la vivienda (Fig. 9).



Fig. 9. Ampliación de la zona de la ciudad en la que se ha realizado la estimación

Finalmente, si se desea, se puede crear un informe como salida más completa (Fig. 10). En él automáticamente aparecen las características de la vivienda y el precio estimado.

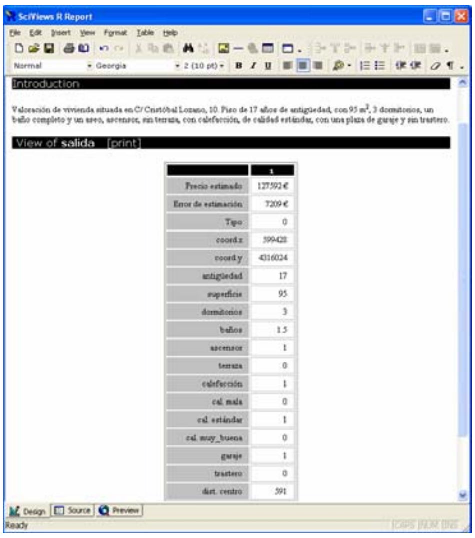


Fig. 10. Informe final de la estimación

5 Conclusiones

En este trabajo se ha demostrado que el uso combinado de las Redes Neuronales Artificiales y los Sistemas de Información Geográfica constituyen una novedosa y potente herramienta en valoración de viviendas en particular y en cualquier problema económico en general donde se vean involucrados datos espaciales.

En cuanto a los resultados particulares de la aplicación, los objetivos se han alcanzado satisfactoriamente. Las redes MLP han resultado más convenientes que los SOFM en la estimación de la calidad de la vivienda. También han sido mejores los resultados alcanzados por el MLP que por las redes RBF en la estimación del precio total. Quizá la razón sea que la muestra (591 registros) no es suficientemente grande para las necesidades de las redes RBF.

Por último, queremos destacar las ventajas de las RNA en la extracción de relaciones no lineales, por otra parte tan frecuentes en Economía.

Lejos de considerar este trabajo como acabado, se intentará en el futuro aprovechar en mayor grado las ventajas del SIG introduciendo más variables de entrada que puedan influir en el precio de la vivienda, como accesibilidad, presencia de zonas verdes, distancia a centros de educación y otros factores socio-económicos y geográficos. Para ello es fundamental el trabajo de actualización de la información geográfica, como también lo es el disponer más fácilmente de datos referidos al mercado inmobiliario para actualizar el sistema automático de valoración y mantener así su utilidad.

Referencias

1. Bishop, C.M.: Neural Networks for Pattern Recognition. Clarendon Press (1995)
2. Do, Q., Grudnitski, G.: A Neural Network Analysis of the Effect of Age on Housing Values. The Journal of Real Estate Research, Vol. 8, n. 2. (1992) 253-264
3. Gámez, M., Montero, J.M., García, N.: Kriging Methodology for Regional Economic Análisis: Estimating the Housing Price in Albacete. International Advances in Economic Research, Vol. 6, n. 3 (2000)
4. García, N.: Diseño de Redes Neuronales Artificiales para el Mercado Inmobiliario. Aplicación a la ciudad de Albacete. Universidad de Castilla-La Mancha. Tesis Doctoral no publicada (2004)
5. Haykin, S.: Neural Nerworks. A Comprehensive Foundation. Prentice Hall (1994)
6. Jacobs, R.A.:Increased rates of Convergente through Learning Rate Adaptation. Neural Networks, vol 1 (4) (1988) 295-307
7. Kohonen, T.: Self-Organizing Maps. 2º ed. Springer-Verlag, Berlín Heidelberg (1997)
8. Martín del Brío, B., Sanz, A.: Redes Neuronales y Sistemas Borrosos. RA-MA (1997)
9. Nguyen, N., Cripps, A.: Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks. The Journal of Real Estate Research, Vol.22, n.3 (2001) 314-326
10. Ripley, B.D.:Pattern Recognition and Neural Networks. Cambridge University Press (1999)
11. Sabella, E.M.: Determining the Relationship Between the Property's Age and Its Market Value. Assesors Journal, 9 (1974) 81-85
12. Thurston, J.: GIS & Artificial Neural Networks: Does Your GIS Think? GISVision Magazine (2002)

Críticos de arte artificiales

Juan Romero[1], Penousal Machado[2], Bill Manaris[3], Antonino Santos[1], Amílcar Cardoso[2] y Marisa Santos[1]

[1] RNASA Lab., Fac. de Informática, Universidade da Coruña, España
{jj, nino, mhyso}@udc.es
[2] CISUC- Centre for Informatics and Systems, Universidade de Coimbra, Portugal
{machado, amilcar}@dei.uc.pt
[3] Computer Science Department, College of Charleston, USA
{manaris@cs.cofc.edu}

Resumen. En este artículo proponemos un marco de trabajo para el desarrollo de críticos de arte artificiales, consistente en una arquitectura y en una metodología de validación. La arquitectura incluye dos módulos: un extractor de características, que lleva a cabo un preprocesamiento de la pieza de arte, extrayendo diversas medidas y características; y un evaluador, que, basado en la salida del extractor de características, clasifica la pieza de arte de acuerdo a un criterio específico. La metodología de validación consta de varias etapas, que van desde la identificación de autor y estilo, hasta la integración del crítico de arte artificial en un entorno dinámico multiagente, que incluye agentes humanos. Usando la estructura propuesta, hemos desarrollado un crítico de arte artificial en el dominio musical, presentado en los resultados experimentales.

1 Introducción

La habilidad para generar piezas de arte es comúnmente asociada con la creatividad. Así, el desarrollo de sistemas computacionales que crean estas piezas pueden contribuir de manera significativa al estudio de la creatividad.

El proceso artístico depende altamente de la habilidad de llevar a cabo juicios estéticos, de inspirarse en el trabajo de otros artistas, y de actuar como crítico del trabajo de uno mismo. Como Boden afirma: “alguien que tiene una nueva idea debe ser capaz de evaluarla por él mismo” [1]. Modelar esta capacidad del artista es un paso importante, sino necesario, en la creación de un artista artificial. Después de todo, un artista es también, y sobre todo, un observador/oyente.

Esto contrasta con la mayoría de los sistemas computacionales que han sido desarrollados durante los últimos años[1]. Típicamente, el rol del observador/oyente ha sido completamente desatendido; tales sistemas no tienen la habilidad de percibir las piezas de arte producidas por ellos (o por otros artistas), ni son capaces de llevar a cabo juicios estéticos. Por lo tanto, estos sistemas tienden a ser completamente cie-

[1] Para una visión general de las aproximaciones computaciones a la composición musical ver, por ejemplo, [2].

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 231-242, 2006.

gos/sordos al mundo exterior.

En este artículo presentamos un marco de trabajo general para el desarrollo de críticos de arte artificiales (CAAs), es decir, sistemas capaces de “ver/escuchar” una obra de arte y llevar a cabo algún tipo de evaluación sobre la misma. Estamos principalmente interesados en CAAs que producen una valoración numérica de las piezas de arte, puesto que esto permite una fácil incorporación del CAA al sistema de generación de piezas de arte. Además, este tipo de valoración no es particular a un estilo de arte específico. Aunque particularmente adecuado a este tipo de CAAs, este marco de trabajo es lo suficientemente general para permitir el desarrollo de otros tipos de CAAs (por ejemplo, aquellos que producen una evaluación descriptiva de la obra de arte).

Este marco de trabajo, basado en un análisis de CAAs existentes, consiste en una *arquitectura* y en una *metodología de validación*.

La arquitectura propuesta consta de un *extractor de características* y un módulo *evaluador*. El extractor de características es responsable de la percepción de la pieza de arte, generando como salida un conjunto de medidas que reflejan sus características relevantes. Estas medidas sirven como entrada al evaluador, que evalúa la pieza de arte según un criterio específico o estético.

Una de las principales dificultades en el desarrollo de artistas computacionales, y más específicamente CAAs, es su validación. Para ayudar a solucionar este problema, proponemos una *metodología de validación multietapa*. La primera etapa de esta metodología permite la valoración objetiva y significativa de los CAAs, proporcionando una base sólida para su desarrollo. Las últimas etapas incorporan criterios más dinámicos, e incluyen probar los CAAs en una sociedad híbrida de humanos y agentes artificiales.

Probamos nuestras ideas mediante el desarrollo de un CAA en el dominio musical, y realizamos un conjunto de experimentos, que, aunque preliminares, dan resultados prometedores.

2 Marco de trabajo para el desarrollo de CAAs

Queremos proporcionar una base para la validación y desarrollo de CAAs que permita la integración de otros críticos y promueva la colaboración entre grupos para la creación de CAAs. El marco de trabajo global está basado en las siguientes características:

- Adaptabilidad – Los CAAs deberían adaptarse a un entorno cambiante. Esto significa replicar una característica comúnmente aceptada entre los críticos humanos: la evolución.
- Sociabilidad – Idealmente, los CAAs deberían poder ajustar su comportamiento según las demandas de la sociedad en la cual están integrados. Esto es, los CAAs deben ser capaces de desarrollarse en un entorno híbrido (un entorno que incorpora humanos y sistemas artificiales). Así, el CAA debe ser validado por una sociedad de “agentes” artificiales y humanos, de la misma forma que los críticos humanos son validados en sociedades puramente humanas [3].

- Generalidad – El CAA debería ser fácilmente adaptable a diferentes dominios; las tareas de dominio específico deberían ser llevadas a cabo por módulos especializados, permitiendo al sistema cambiar fácilmente de un dominio a otro.
- Independencia de representación – El CAA debería construir su propia representación interna de la pieza de arte, determinando su evaluación a partir de la misma; y solo debería tener acceso a la pieza de arte, no a otro tipo de nivel superior de representación de la misma.

2.1 Arquitectura

Teniendo en cuenta las características presentadas previamente, la arquitectura debe permitir el desarrollo de CAAs fácilmente adaptables a diferentes dominios, y que consideran las particularidades de los mismos. Por ejemplo, la forma de tratar con la música y con el arte visual es visiblemente distinta: mientras que la música sigue una secuencia temporal predeterminada, el arte se observa mediante un acceso más directo a la pieza arte. De ahí la necesidad de dividir el sistema en varios módulos, siendo específicos en tareas particulares a un dominio, y permitiendo la generalidad de los módulos restantes.

Normalmente las piezas de arte contienen una gran cantidad de información. En arte visual, por ejemplo, incluso un cuadro relativamente pequeño consume una gran cantidad de memoria. Como se puede deducir a partir del análisis del estado del arte de sistemas adaptativos actuales (redes de neuronas, algoritmos genéticos...), tales cantidades de información no pueden ser manejadas razonablemente. Para abordar este problema, algunos investigadores recurren a la reducción del tamaño de las obras de arte que alimentan el sistema adaptativo (p.ej. [4]). Sin embargo, esta aproximación implica una importante pérdida de información y detalle, y los resultados experimentales son, típicamente, decepcionantes.

Creemos que hay una aproximación más adecuada, que consiste en preprocesar las piezas para extraer características relevantes, que pueden entonces ser usadas como entrada a la parte adaptativa del sistema. Esto reduce la cantidad de información que tiene que ser procesada.

La arquitectura propuesta incluye dos módulos: el *extractor de características* y el *evaluador*. Cada módulo tiene un propósito concreto y diferente. El extractor de características lleva a cabo un análisis de la pieza de arte y proporciona un conjunto de características relevantes al evaluador. El evaluador realiza una valoración de la pieza basado en las características extraídas previamente.

El extractor de características lleva a cabo dos tareas específicas: percepción y análisis. Durante la percepción, el sistema construye un tipo de representación interna de la pieza de arte. Después, en la tarea de análisis, esta representación es analizada y proporciona un conjunto de medidas relevantes. Mientras que esta separación entre percepción y análisis es principalmente conceptual, la idea es que, en una primera etapa, el extractor de características adquiera información sobre parámetros específicos al dominio que son analizados posteriormente.

La representación interna no está restringida a técnicas específicas, sino que puede ser de diferentes tipos: estadística, algorítmica, simbólica, no simbólica, etc.

El evaluador es un sistema adaptativo que toma como entrada la caracterización de la pieza de arte realizada por el extractor de características, y obtiene como salida una valoración de dicha pieza.

Este módulo debe adaptarse a diferentes tareas según la información de retroalimentación que se le proporciona. Dependiendo de la tarea, esta información indica la respuesta deseada o una evaluación de la actuación del CAA, el cual debe ajustar su comportamiento para maximizar su rendimiento.

Además, el módulo evaluador adaptativo puede dar información acerca de las características que son relevantes en la evaluación de una pieza de arte. Los pesos de una red de neuronas artificiales (RNA), por ejemplo, pueden mostrar qué características son las más significativas a la hora de criticar una pieza. También es posible encontrar el conjunto mínimo de características necesarias para una cierta tarea mediante pruebas de test al evaluador con diferentes conjuntos.

La arquitectura propuesta permite que la búsqueda de características relevantes y la evaluación sean independendientes. Así, es posible incluír extractores de características y evaluadores de diferentes investigadores que permitan probar qué combinación de extractor y evaluador es la que caracteriza mejor a la pieza de arte. Ahora presentaremos una metodología de validación que fue diseñada para realizar un test estructurado del CAA desarrollado.

2.2 Metodología de validación

La validación de un CAA presenta principalmente dos dificultades: la subjetividad existente en la evaluación de las piezas de arte, y el hecho de que sean necesarios grandes conjuntos de entrenamiento para entrenar el módulo evaluador (cientos de piezas de arte evaluadas por humanos).

La respuesta a estas complicaciones es el uso de una *metodología de validación multietapa*. En cada nivel, el CAA se presenta con una tarea diferente. Se empieza con tareas en las que la exactitud de la salida puede ser objetivamente determinada, y que no requieren un conjunto de piezas de arte evaluadas por humanos. En la siguiente etapa las tareas requieren una mayor subjetividad y complejidad. En los primeros niveles la respuesta del CAA es estática; en el último nivel, sin embargo, el CAA debe adaptarse al entorno y cambiar su evaluación en el tiempo según el contexto que lo rodea.

Actualmente, consideramos tres niveles de validación: Identificación, Evaluación Estática y Evaluación Dinámica.

El *nivel de identificación* se ocupa de la valoración de la habilidad del CAA para reconocer el autor o el estilo de una determinada obra.

En la tarea de *identificación de autor* se presenta al CAA varias piezas de arte de diferentes autores. Su objetivo es determinar el autor de cada pieza. El módulo evaluador es entrenado con información de retroalimentación que le indica la tarea a realizar. Este tipo de validación es relativamente fácil de llevar a cabo, la compilación de instancias de entrenamiento es sencilla, y el test es totalmente objetivo. La principal dificultad que afecta a este nivel de prueba es la construcción de conjuntos de entrenamiento y de test representativos.

La tarea de *identificación de estilo* es similar a la anterior. La diferencia es que en este caso el CAA debe identificar el estilo de una obra. El entrenamiento y el test pueden ser realizados del mismo modo que en la etapa de identificación de autor. Este tipo de validación permite la comprobación de CAAs que pueden ser usados en una amplia variedad de tareas, tales como recuperación de imágenes y música, permitiendo búsquedas basadas en el estilo.

La principal dificultad de estas tareas depende de los artistas y estilos escogidos. Intentar discriminar entre artistas de la misma escuela puede ser más difícil que distinguir estilos radicalmente distintos. Sin embargo, discriminar entre artistas que tienen rasgos característicos (en el sentido usado en [5]) es más fácil que entre estilos relacionados. En el análisis de los resultados de los experimentos es importante tener en cuenta qué es razonable esperar. Por ejemplo, si el conjunto de test incluye obras atípicas, el CAA probablemente fallará. Esto no indica necesariamente un defecto del extractor de características o del evaluador, sino simplemente el hecho de que una pieza de arte es atípica.

Aunque de ámbito limitado, las tareas de identificación son útiles para determinar las capacidades del módulo extractor de características. Un fallo en estas pruebas puede indicar que el conjunto de características extraídas no es suficiente para discriminar entre autores o estilos, previniéndonos de esta manera desplazarnos a una tarea más compleja, que seguramente falle debido a la falta de información significativa. Además, un análisis de las características usadas por el evaluador en las tareas de identificación puede ayudar a determinar la importancia relativa de cada una de ellas. De hecho, de se pueden desarrollar tests específicos para conocer el poder predictivo de cada medida o conjunto de medidas.

El segundo nivel de validación es la evaluación estática. El objetivo del CAA es determinar el valor estético de una serie de piezas de arte previamente evaluadas por humanos. Una de las mayores dificultades para desarrollar este test es la construcción de una base de datos representativa con obras evaluadas consistentemente.

Es importante darse cuenta de que el entrenamiento del CAA requiere no solo ejemplos positivos, sino también negativos. Irónicamente, es bastante difícil conseguir un conjunto representativo de “cosas incorrectas que hacer”.

Para crear el conjunto de entrenamiento, se puede recurrir a una herramienta de arte generativa. Esto produciría un alto número de piezas en una cantidad de tiempo razonable. Sin embargo, la consistencia de la evaluación depende mayormente de la disciplina del usuario. Adicionalmente, el conjunto será solo representativo de piezas típicamente creadas por una herramienta de arte generativa. Además, el grado de correlación entre las piezas creadas puede ser alto, haciendo la tarea del CAA más fácil.

Otra opción sería disminuir el ámbito de aplicación del CAA; esto es, crear un CAA que es capaz de evaluar la calidad estética dentro de un estilo bien definido. Esto da lugar a un paso de validación que está de alguna forma más cercano a la tarea de identificación de estilo y que es, por tanto, menos subjetiva. La diferencia está en que el CAA está valorando la distancia a un estilo dado en vez de intentar distinguir entre estilos.

El análisis de los resultados experimentales puede ser muy significativo; uno necesita estar seguro de que el CAA está llevando a cabo la tarea esperada y no explotan-

do algún defecto del conjunto de entrenamiento. Por ejemplo en [6] los autores entrenaban un sistema de reconocimiento de caras, que tenía un resultado sorprendentemente bueno. Sin embargo, un cuidadoso análisis de los resultados mostró que el sistema no estaba reconociendo las caras de las personas que aparecían en las imágenes, sino las oficinas en las cuales estaban tomadas las fotos.

La etapa de evaluación estática tiene muchas dificultades, tanto en la construcción del test como en el análisis de los resultados experimentales, pero es, sin embargo, necesaria para poder evaluar un CAA.

El último paso en la metodología es la *evaluación dinámica*. El valor de una obra depende del contexto cultural que la rodea. Así, el CAA debe ser "consciente" de este contexto, y ser capaz de adaptar su evaluación a los cambios que se dan en su entorno. Esto es, su comportamiento debe ser socialmente adecuado. Para desarrollar esta validación, se propone un modelo de sociedad llamado "Sociedad Híbrida" (SH). SH es un paradigma similar a la Vida Artificial, pero con "agentes" humanos en el mismo nivel que los artificiales. La sociedad híbrida explora la creación de sociedades igualitarias pobladas por seres humanos y artificiales en dominios artísticos (o sociales); como tal, SH es adecuada para validar el CAA de una forma natural y dinámica. En esta etapa, el éxito del CAA depende de la estimación de sus juicios por los otros miembros de la sociedad. Este tipo de test introduce una nueva dimensión social y dinámica a la validación, puesto que el valor de una obra varía con el tiempo, y depende de los agentes que componen la sociedad.

El problema de este nivel de validación es la necesidad de incorporar humanos en la experimentación. Así, los experimentos son difíciles de organizar, y existen limitaciones de tiempo. Además, la capacidad de adaptación de los críticos debe ser alta para poder adaptarse a un entorno dinámico y complejo. A pesar de las dificultades inherentes, estos críticos pueden ser evaluados y fácilmente integrados en la "sociedad de la información" como asistentes de usuarios o como parte de sistemas generales.

En los dos primeros niveles de validación es posible valorar la ejecución del feature extractor y del evaluador independientemente, puesto que la salida del feature extractor (junto con la información de retroalimentación), puede ser vista como una instancia de entrenamiento para el evaluador. En el tercer nivel, esto ya no es posible ya que la información de retroalimentación no refleja directamente la calidad de las obras, sino solamente una estimación de las acciones del CAA mediante la sociedad, que cambia dinámicamente con el tiempo.

La metodología de validación aquí presentada intenta encontrar un compromiso entre la validación humana y la automática. Conscientes de la dificultad de las tareas propuestas, es importante resaltar que para ciertas tareas solo se necesitan tener en cuenta algunos de los niveles de validación.

3 Experimentos: Resultados

Usando el marco de trabajo presentado, desarrollamos un CAA en el dominio musical y comprobamos su comportamiento mediante un conjunto de experimentos, que se corresponden con el primer nivel de validación. La tarea presentada al CAA es dis-

criminar entre obras musicales de dos compositores: Beethoven y Bach. El conjunto de piezas musicales consiste en 108 partituras de Bach (una colección de sonatas, preludios, fugas, fantasías, toccatas, conciertos, etc.) y 32 sonatas de piano de Beethoven.

Siguiendo la arquitectura propuesta, el sistema tiene dos módulos, el extractor de características y el evaluador adaptativo, que son descritos en las siguientes secciones.

3.1 Extractor de características

El extractor de características, descrito en [7], utiliza una colección de métricas, basadas en la Ley de Zipf [8], para extraer una serie de métricas a partir de piezas de músicas codificadas en formato MIDI.

Las distribuciones Zipf han sido descubiertas en un amplio rango de fenómenos, incluyendo la música. Por ejemplo, en [9] se presenta un estudio de 220 piezas de varios estilos de música (barroco, clásico, romántico, doce-tonos, jazz, rock, cadenas de ADN, y música aleatoria), descubriendo en ellas varias distribuciones Zipf.

En los experimentos se usan un total de 40 métricas, además del número de notas de la obra. Cada una de las 40 métricas produce dos números reales:

1. La *pendiente* de la línea de dirección de las frecuencias de evento, trazada en un formato log-log, rango-frecuencia; este número varía entre 0 y $-\infty$, con -1 denotando una distribución Zipf; y
2. La fuerza de la correlación lineal, R^2, de la línea de dirección; ésta abarca desde el 0 al 1, siendo el 1 el que indica un ajuste perfecto.

Las métricas usadas en el extractor de características se pueden dividir en 3 tipos:

- Las *métricas globales* proporcionan información estadística de la pieza como un todo. Hay siete métricas de este tipo: tono, tono-relativo-a-octava, duración×tono, duración×tono-relativo-a-octava, intervalo melódico, intervalo harmónico e intervalo melódico-harmónico.
- Las *métricas estructurales* miden el equilibrio de órdenes más altos de cambio de tono. Actualmente, capturamos seis órdenes de cambio. Las métricas de primer-orden miden el equilibrio de los cambios en los intervalos melódicos. Las métricas de segundo-orden miden el equilibrio de los cambios entre los intervalos de primer-orden, y así el resto.
- Las *métricas fractales* miden la dimensión fractal de cada una de las métricas anteriores. Estas métricas aplican recursivamente una métrica dada en diferentes niveles de resolución dentro de una pieza. Mediante la subdivisión sucesiva de la pieza en partes, la carencia de equilibrio local puede ser expuesto. Como las otras métricas, las métricas fractales producen una pendiente y un valor del error cuadrático medio. La pendiente es equivalente a la dimensión fractal de la métrica dada. El proceso de particionamiento para cuando alcanzamos frases con menos de cinco notas.

3.2 Evaluador adaptativo

El evaluador adaptativo usado en los experimentos consiste en una *RNA feed-forward* (con alimentación hacia delante) con una capa oculta. Tras probar con diferentes arquitecturas de redes, elegimos una con 30 neuronas en la capa de entrada, 12 en la oculta y 2 en la de salida. Cada unidad de la capa de entrada se corresponde con cada uno de los valores generados por las métricas. Estos valores se normalizan en el intervalo [-1, 1]. Una salida (1, 0) indica que el autor de la partitura es Beethoven, mientras que (0, 1) indica que pertenece a Bach. El conjunto de entrenamiento usado en los experimentos contiene un 66% de las partituras (aleatoriamente seleccionadas) de cada compositor. El conjunto de test contiene las restantes.

En los experimentos preliminares, se descubrió que cuando no se incluía una partitura atípica en el conjunto de entrenamiento, la RNA fallaba al identificar a su autor. Este problema se solucionó al incluir esta partitura en el conjunto de entrenamiento. En la sección 3.3 se hace un análisis detallado de este problema.

Usamos el SNNS2 para construir, entrenar y probar las RNAs. Los pesos son aleatoriamente inicializados con valores del intervalo [-1, 1]. La función de aprendizaje utilizada es la back-propagation estándar (propagación hacia atrás), con una tasa de aprendizaje de 0.1 y un momentum igual a 0. En los primeros tests usamos una RNA completamente conectada. El entrenamiento de la RNA se realiza con 30000 ciclos. Tras el entrenamiento la red es capaz de identificar correctamente todas las partituras de los conjuntos de entrenamiento y test. El error cuadrático medio (ECM) en el ciclo 30000 era 0.00003 en el conjunto de entrenamiento, y 0.00576 en el conjunto de test.

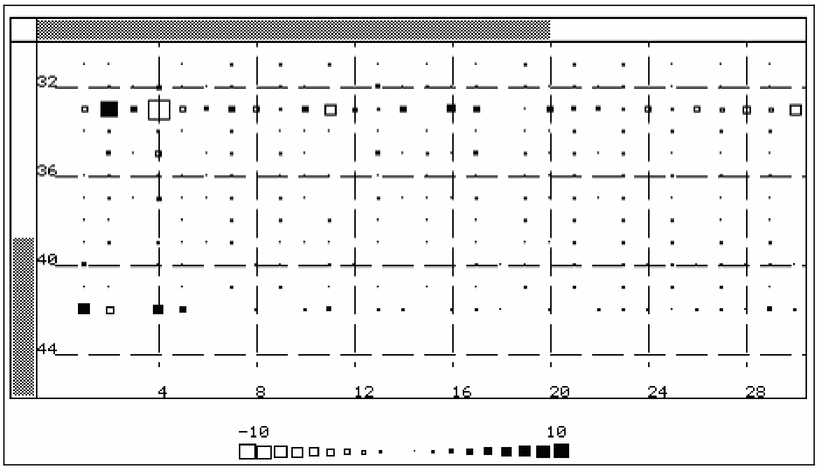


Fig. 1. Conexión entre los pesos de los elementos de procesamiento de la capa de entrada (eje x) y los de la capa oculta (eje y) de la RNA

[2] Stuttgart Neural Network Simulator (http://www-ra.informatik.uni-tuebingen.de/SNNS/).

Para discriminar las características más relevantes para la tarea de identificación, sumamos los pesos en valor absoluto de las conexiones de cada una de las neuronas de la capa de entrada con las neuronas de la capa oculta (Figure 1). Los valores más altos están asociados a las características más relevantes. Tras varias repeticiones del experimento, se observan similares distribuciones de pesos vinculados a las mismas características. Para comprobar esto se realizaron los siguientes tests.

Comenzamos sacando las ocho neuronas menos significativas. Los resultados obtenidos son similares a los anteriores. Tras 30000 ciclos la red identifica correctamente todas las partituras de los conjuntos de entrenamiento y test (ECMs de 0.00010 y 0.00356, respectivamente).

Después construimos una RNA con sólo 6 unidades de entrada, correspondientes con las características más relevantes. Los resultados muestran una ligera degradación del proceso. En el ciclo 10000, la RNA identifica correctamente todas las partituras del conjunto de entrenamiento y el 94% de las del conjunto de test (ECMs de 0.00391 y 0.11166, respectivamente). A partir de este punto, el error en el test se incrementa graducalmente, alcanzando un ECM de 0.12776 en el ciclo 70000, lo que indica que la RNA se ha sobreentrenado.

3.3 Análisis de los experimentos

Los resultados obtenidos muestran que las métricas basadas en Zipf combinadas con la RNA, son suficientes para la identificación de autor. Además se destaca un conjunto de seis características relevantes para discriminar entre los dos autores. Cabe mencionar que esto no significa necesariamente que sean las características más importantes en diferentes tareas o con diferentes autores.

Las Figuras 2.a y 2.b muestran los mapas de contorno en 3D de las obras de Bach y Beethoven. En estas visualizaciones, el eje x se corresponde con las 6 características más significativas; el eje y se corresponde con una pieza musical (1 a 32); y el eje z se corresponde con el valor absoluto de la característica. Analizando estas figuras se ve que la pendiente del tono-relativo-a-octava es uno de los factores clave para discernir las obras de Bach y Beethoven. Como se muestra, las piezas de Bach exhiben una distribución casi-Zipfiana de los tonos de la escala-12-cromática (la pendiente media es -1.1629; std 0.2809), mientras que las sonatas de Beethoven tienden a ser más distribuidas uniformemente (pendiente media -0.8343; std 0.2188).

La Figura 2.b permite la identificación de una pieza, la Sonata para Piano no. 20 de Beethoven, que no se ajusta al contorno de Beethoven. La pendiente del tono-relativo-a-octava para esta partitura es -1.7472. Esta pieza es atípica con respecto a las otras obras de Beethoven y a las características usadas. Como tal, para ser clasificada correctamente, esta pieza tuvo que ser incluída en el conjunto de entrenamiento. La inclusión de esta instancia de entrenamiento fuerza a la RNA a basar su valoración en un conjunto más amplio de características, lo que, aunque provoca un entrenamiento más lento, fomenta la robustez y generalización de la RNA.

Una característica secundaria distintiva entre los dos contornos musicales es el R^2 del tono-relativo-a-octava. Las obras de Bach tienden a producir una línea de dirección más dispersa (media R^2 0.6612; std 0.0787), mientras que las de Beethoven

tienden a tener una línea de dirección más ajustada (media R^2 0.8017; std 0.0571). Aunque éstas son las dos características más destacadas, es necesarion tener en cuenta el conjunto de las seis características para lograr la correcta identificación de todas las partituras.

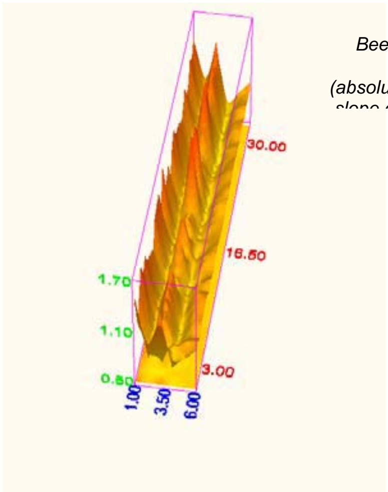


Fig. 3.a. Bach- Mapa de contorno de 6 características sobre 32 piezas de Bach (BWV 500-531)

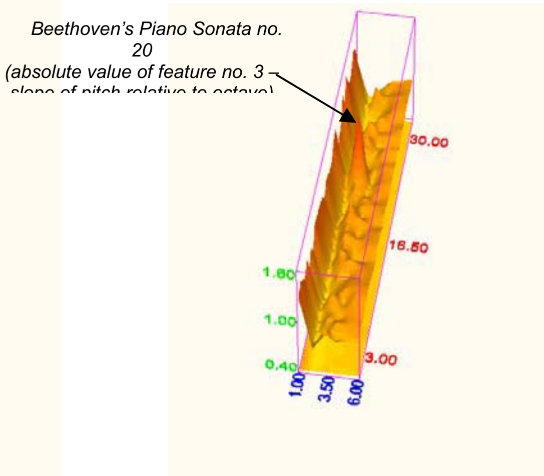


Fig. 3.b. Beethoven- Mapa de contorno de 6 características sobre 32 piezas de Beethoven (piano sonatas 1-32)

4 Conclusiones y futuros trabajos

Proponemos un marco de trabajo genérico para el desarrollo de críticos de arte artificiales, basado en el análisis del estado del arte actual en el área, y en la experiencia adquirida en el desarrollo de sistemas previos. Este marco de trabajo incluye una arquitectura y una metodología de validación. Para permitir una fácil adaptación a diferentes dominios, la arquitectura propuesta separa los componentes genéricos del dominio de los específicos. Además también establece un límite entre los módulos estáticos y adaptativos. La validación de CAAs es una tarea compleja, por lo que la metodología multinivel permite probarlos estructuradamente y compararlos con diferentes aproximaciones.

Siguiendo con el marco propuesto, implementamos un CAA y probamos su eficiencia en la identificación de autor. El extractor de características obtiene una serie de métricas basadas en Zipf, que sirven como entrada al evaluador adaptativo, implementado mediante una RNA. Una vez entrenada es capaz de reconocer todas las instancias de los conjuntos de entrenamiento y test, mostrando la eficacia de las métricas basadas en Zipf para la identificación de autor. Éste es el primer uso de dichas métricas para la atribución de autoría en música. Un análisis de la RNA nos permitió identificar el conjunto de características que son más importantes para la discrimina-

ción de autores. La identificación de estas características puede ser útil desde una perspectiva musical, dando una idea de la caracterización de los estilos de los autores. La naturaleza modular de la arquitectura permite una fácil: integración de características adicionales y la adaptación a otros dominios.

Actualmente estamos desarrollando numerosos experimentos en el dominio del arte visual y musical, que incluye la discriminación entre más autores, con un conjunto de características mayor. Una posibilidad interesante consiste en explorar si el principio de mínimo esfuerzo de Zipf podría ser usado, en un nivel más alto, para evaluar la eficacia y la "naturalidad" de una sociedad arbitraria igualitaria de CAAs, mediante el examen de diversos aspectos de la interacción social entre agentes.

El marco de trabajo aquí descrito no está restringido a dominios artísticos. Puede ser usado en algunos dominios que incluyen (a) la creación de una hipótesis (diseño, solución, etc.), y (b) el refinamiento iterativo de aquellas hipótesis basadas en la estéticas, restricciones, y otros atributos cuantificables. Tales dominios incluyen el desarrollo de software, matemáticas, ingeniería, y arquitectura. Por ejemplo, las métricas Zipf ya han sido usadas para evaluar software, diseño arquitectónico y otros sistemas complejos [10; 11]. Adicionalmente, la aplicación a otras áreas tales como imágenes basadas en contenido, y búsqueda y recuperación de música, también parece viable.

La investigación en el área de los críticos de arte artificiales todavía está en una etapa embriónica. El sistema propuesto pretende proporcionar una base común para el desarrollo y la validación de críticos de arte artificiales, y promocionar la colaboración entre investigadores en esta área.

Agradecimientos

Expresamos nuestra gratitud a Robert Davis por sus diversos análisis estadísticos; Charles McCOrmick, Tarsem Purewal, Dallas Vaughan y Christopher Wagner por contribuir en el desarrollo de las métricas basadas en Zipf.

Referencias

1. M. A. Boden. The Creative Mind: Myths and Mechanisms. London, Cardinal. 1990.
2. G. Papadopoulos and and G. A. Wiggins. AI Methods for Algorithmic Composition: A Survey, A Critical View, and Future Prospects. *Proceedings of the AISB'99 Symposium on Musical Creativity*, 1999.
3. A. Pazos, A. Santos, B. Arcay, J. Dorado, J: Romero, and J. Rodríguez. An Application Framework for Building Evolutionary Computer Systems in Music. *Leonardo*, 36(1), 2003.
4. S. Baluja, D. Pomerleau, and T. Jochem. Towards Automated Artificial Evolution for Computer-Generated Images. In *Connection Science* 6, No. 2, pp. 325–354. 1994.
5. David Cope. *Experiments in Musical Intelligence*. Madison, WI: A-R Editions, 1996.
6. A. Teller and M. Veloso. Algorithm evolution for face recognition: What makes a picture difficult. *In Proceedings of the International Conference on Evolutionary Computation*, IEEE Press, 1995.
7. B. Manaris, T. Purewal and C. McCormick. Progress Towards Recognizing and Classifying

Beautiful Music with Computers-MIDI-Encoded Music and the Zipf-Mandelbrot Law. In *Proceedings of EEE SoutheastCon 2002*, Columbia, SC, pp. 52–57, 2002.
8. G. K. Zipf. *Human Behavior and the Principle of Least Effort*. New York: Hafner Publishing Company, 1949.
9. Manaris, B., Vaughan, D., Wagner, C., Romero, J., and Davis, R.: Evolutionary Music and the Zipf-Mandelbrot Law: Developing Fitness Functions for Pleasant Music. In: Lecture Notes in Computer Science, Applications of Evolutionary Computing. LNCS 2611, Springer-Verlag, pp. 522–534, 2003.
10. M. Shooman and A. Laemmel. Statistical Theory of Computer Programs. In *Proceedings of IEEE Computer Conference*, pp. 511–517, Oct. 1977.
11. N.A. Salingaros and B. J.West. A Universal Rule for the Distribution of Sizes. *Environment and Planning,* B(26), pp. 909–923, 1999.

TOPOS: Reconocimiento de patrones temporales en sonidos reales con redes neuronales de pulsos

Pablo González Nalda[1] y Blanca Cases[2]

[1]Escuela Universitaria de Ingeniería de Vitoria-Gasteiz, UPV-EHU
pablo@si.ehu.es http://lsi.vc.ehu.es/pablogn/
[2]Facultad de Informática de Donostia-San Sebastián, UPV-EHU

Resumen En este artículo hacemos un resumen de la aplicación TOPOS y presentamos nuevos resultados obtenidos con la misma.
El sistema desarrolla un robot que visita en un determinado orden dos fuentes de sonidos reales. Los dos sonidos emiten exactamente el mismo sonido del canto de un canario, con la única diferencia de que cada sonido tiene un orden distinto en sus partes, parecido a "per" y "pre".
Para diferenciar dos señales que se reciben simultáneamente el robot se beneficia de las características de situación y corporeidad, y de las ventajas que ofrecen las redes neuronales de pulsos (PCNN).
TOPOS es un modelo en el que se evolucionan poblaciones de robots del tipo Khepera dentro de un esquema evolutivo con un fuerte referente biológico, con el fin de obtener comportamientos de navegación a través del reconocimiento de señales complejas y variables en el tiempo. Se modelan las estructuras que llevan a cabo la percepción auditiva, entre ellas la cóclea mediante la Transformada de Fourier.

1. Introducción. Vehículos de Braitenberg y robótica evolutiva

En este artículo se presenta un resumen de los fundamentos que permiten comprender los objetivos y la estructura y mecanismos de la aplicación TOPOS[1]. Por otra parte, describimos nuevos resultados obtenidos con la misma, resultados que dan una muestra de las posibilidades que nos proporciona el planteamiento en el que se ha desarrollado.

Los vehículos de Braitenberg [2] son experimentos mentales que implementan tropismos y taxias. El vehículo de Braitenberg más interesante se dirige a una fuente que estimula los sensores, por ejemplo una fuente de luz y sensores basados en células fotoeléctricas. Se denomina fototropismo positivo, y sería un simple mecanismo cableado (*hardwired*) simétrico en el que un aumento del estímulo en la célula izquierda aceleraría el motor derecho y viceversa, con lo que el vehículo se mueve hacia la fuente de luz.

Este trabajo está encuadrado en la Robótica Autónoma [3] y en su rama darwiniana, la Robótica Evolutiva [4]. En ésta se evolucionan poblaciones de

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 243-248, 2006.

robots probándolos con una determinada tarea para después seleccionarlos según la puntuación que hayan obtenido. Los objetos de evolución pueden ser modelos de robots, simulaciones de robots o los probios robots reales.

La aplicación presentada en [1] se denomina TOPOS por el problema de navegación basado en el reconocimiento de señales complejas que actúan como puntos de referencia o *landmarks* en un entorno no estructurado. En concreto, las señales son sonidos como los que se pueden grabar directamente en la naturaleza, por ejemplo el canto de un pájaro. La percepción está fuertemente ligada al movimiento de la cabeza y las orejas para deducir el origen del sonido y usarlo para orientarse.

En estas ramas de la Robótica hay trabajos que desarrollan vehículos de Braitenberg que usan luz blanca en robots del tipo Khepera [5]. Hay también algunos diseños que usan sonido en vez de luz, como los que emulan la fonotaxia del grillo en robots que reconocen el canto de un grillo de una especie determinada, cuatro ráfagas de 20 ms de una onda de 4'8 kHz [6]. La mayor parte de los trabajos desarrollados hasta el momento usan señales simples y constantes como estímulos. Sin embargo, algunos artículos presentan avances, como en [7], que se usa una distancia a un objeto para distinguir cuál de los dos posibles es, y estudian la capacidad de distinguir secuencias de hasta tres bits. En [8] subrayan la dificultad de reconocer señales complejas que varían en el tiempo, y preparan un Khepera que permite que no choque con paredes pintadas con franjas verticales, que parecen código de barras.

2. Sonido y sensores: Fourier y espectro de frecuencias

El sonido lo percibimos [9][10] con la unión del sistema auditivo y del nervioso, procesando la señal mediante la forma del oído externo, la oreja (ecolocación), el tímpano y los huesecillos del oído medio (compresión del rango dinámico), y la estructura neuronal que recibe los movimientos ciliares de la cóclea.

La señal que .ºyen"los *topos* es la parte real de la Transformada de Fourier, obtenida de cada 0'04 segundos de sonido, y que se representa como un vector de 64 números reales. Cada valor representa la intensidad de un sonido en un cierto rango de frecuencias o *bandas*. Cada una de las 64 bandas tiene una anchura de 47 Hz, y por tanto, la frecuencia máxima representada es de 3000 Hz (ver la figura 1). Los sensores se activan más intensamente en una frecuencia, llamada característica (FC), cuyo umbral es el más bajo. Las bandas adyacentes a la banda de esa frecuencia producen menor activación. Cada sensor está definido por su FC, umbral y nivel de saturación. De esta manera se selecciona, de forma adaptativa, la información útil para reconocer la señal.

3. Aplicación TOPOS

TOPOS modela el clásico experimento de la *caja de Skinner*[11]. Una rata en una caja debe (o puede) pulsar una palanca u otra. Las ratas aprenden a relacionar

el premio o castigo que reciben al apretar una palanca con el estímulo asociado a la palanca (una luz, una imagen, un sonido).

TOPOS [1] es un modelo, es decir, no pretende reflejar con precisión aspectos biológicos (aunque tiene un fuerte referente biológico) ni la dinámica o forma de un robot en concreto. Este modelo es un paso importante hacia la obtención de un controlador para un robot del tipo Khepera que desarrolle la tarea, ya que en este modelo no se imponen fuertes restricciones y en [12] se justifica que es factible la conversión de un modelo relativamente simple a un robot.

El agente está situado y corporeizado (integrado en el entorno y en su propio cuerpo, *situated and embodied*): los sensores tienen una determinada forma que modifica la señal que pasan a la red neuronal, y la posición relativa del individuo respecto a la de las fuentes de sonido determina la intensidad de la señal. El diseño a mano es casi imposible [13], así que se necesita evolucionar el controlador (una red neuronal) de forma integrada con los sensores y motores y con el entorno [14]. Así, los individuos (que llamamos *topos*) se adaptan filogenéticamente.

En TOPOS se hace una generalización de la estructura de un vehiculo de Braitenberg al usar una red neuronal formada por dos subredes simétricas parcialmente interconectadas (comparten 4 neuronas). Cada una de las dos subredes es una red de 8 neuronas, de tipo PCNN, es decir, redes neuronales recurrentes totalmente conectadas y de pulsos con retardos en los axones [15]. La red tiene 12 neuronas en total y 12 sensores (cada uno conectado a una neurona, ya sea a la subred del mismo lado o al del otro). En este esquema las neuronas no calculan un valor de salida aplicando una función sigmoidea a la suma ponderada de las entradas, sino que se disparan (y por tanto, tienen dos estados, en proceso de disparo o en reposo) cuando se supera el umbral.

Varios trabajos en Robótica Evolutiva [16][4] usan *Continuous Time Recurrent Neural Networks (CTRNN)*, que se pueden describir fácilmente como un sistema de ecuaciones diferenciales. En los dos modelos se usa un modelo recurrente (con ciclos), pero en cambio, en las PCNN las neuronas disparan pulsos simulando la mayor parte de neuronas naturales [17]. Las ventajas de estas PCNN es que son biológicamente plausibles, integran percepciones en el tiempo, procesan información temporal con los retardos en los axones, y que son más resistentes al ruido, equivalentes a las sigmoideas y a veces con menos neuronas[15].

TOPOS es un programa de ordenador en el que se desarrolla una estructura de un sistema de Robótica Evolutiva típico, en el que encontramos una población de un cierto número de robots (en este caso, modelos idealizados de robots del tipo Khepera), cuyos parámetros se han obtenido de su genoma, y que son evaluados en una tarea, seleccionados, y cruzados los mejores para obtener una nueva generación.

El genoma es un vector de números reales que determinan los valores de sensores (frecuencia, umbral y nivel de saturación, sensibilidad a frecuencias cercanas), neuronas, retardos en axones, pesos sinápticos y velocidad de los motores. El Algoritmo Genético y la selección mantienen una *élite* de los mejores topos de cada generación (25 %) y el resto de la población se genera por cruce de dos progenitores (excluyendo a los 25 % peores de la generación anterior).

La tarea consiste en visitar en un orden determinado al principio de la ejecución dos puntos en un plano, colocando al comienzo el robot en un lugar equidistante de ambos (10'5 unidades). En cada punto están situadas (asignadas aleatoriamente) dos fuentes de sonido. Por visitar se entiende acercarse a una distancia determinada a la fuente (en el experimento es 5'25 unidades, la distancia entre los faros dividida entre 4). El robot, según lo que reciba por sus sensores podrá moverse por el plano activando con diferente velociadad los motores. La función de adecuación (*fitness*) se define a partir de la puntuación de cada una de las pruebas hechas a cada *topo* de la población. La puntuación de una prueba depende de la distancia mínima a cada uno de los dos*faros* o fuentes de sonido. Se restan los valores de las distancias mínimas a una cantidad base de 100. Se suma un *bonus A* (10+var) si el topo visita la primera fuente en primer lugar. El valor variable es mayor cuanto menor sea la distancia al segundo faro. Se suma un segundo bonus *bonus B* de valor 30 si además se visita el segundo faro en segunda posición. La mayor puntuación corresponde con la mejor actuación. En caso de hacer por error una mala prueba, la puntuación sería negativa, pero se pone a 0 para que no pese en exceso en la nota global. La puntuación es un valor absoluto, no depende del entorno ni del comportamiento de otros topos. La máxima puntuación será $100 - 5{,}25 - 5{,}25 + 10 + (21 - 5{,}25) + 30 = 129{,}5$

4. Resultados del experimento

Se ha llevado a cabo un experimento con una población de 100 topos, y cinco pruebas de 35 segundos por individuo y generación. La última generación (200) es una sesión especial de 100 pruebas a cada individuo de la élite para medir la capacidad de respuesta al problema y así obtener datos estadísticos de los mejores individuos de la ejecución. En vez de usar la puntuación, contamos cuántas veces se realiza la tarea correctamente (*acierto*), en cuántas se falla, y el resto son las veces en las que no es capaz de "decidirse". Con estos datos definimos estos valores:

- eficacia relativa `efr` = aciertos/(aciertos+fallos)
- eficacia absoluta `efa` = aciertos/pruebas

Los sonidos son iguales excepto en el orden de las partes (ver la figura 1). Podría compararse a diferenciar entre "perz "pre.ᴬdemás debe visitar los dos faros en un orden determinado para toda la ejecución para obtener un *acierto*. Se determinaría *fallo* ir primero al faro que debe visitar en segundo lugar.

El canto del pájaro de cada uno de los dos sonidos tiene una parte de silencio antes o después del sonido, como se ve en la figura 1, pero como el punto de comienzo de la reproducción del sonido es aleatorio en cada prueba, a veces serán simultáneos y más difíciles, por ejemplo uno parecerá un extraño eco del otro, o eliminará la posibilidad de oír cierta parte clave del sonido.

Denominamos los experimentos *AB*, *BA* y *AA* (tabla 1). La primera letra indica el sonido que se debe visitar primero. La doble A significa que damos el mismo sonido, con el mismo orden en sus partes, tanto al faro al que se

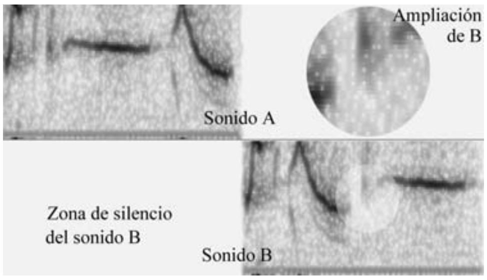


Figura 1. Espectro de frecuencias del canto de un pájaro. El tiempo es el eje horizontal (segundos), la frecuencia en Hz (vertical), y las mayores amplitudes en negro (escala de B/N logarítmica). La altura de cada cuadrado es la anchura de una banda

debe visitar primero como al segundo. En esta situación no hay información para escoger ninguno, por tanto la mejor estrategia parece ser acercarse pero no *marcar*, no jugar.

Exp	EFR $\overline{x}$	EFR σ	EFA $\overline{x}$	EFA σ
AB	100	0	20'8	4'80
BA	100	0	19'8	5'61
AA	4'0	19'6	0'2	1'19

Cuadro 1. Media $\overline{x}$ y desviación estándar σ de EFR y EFA de los tres experimentos.

En resumen, parece que una de cada cinco veces tiene la información necesaria para decidirse a realizar la tarea, que la efectúa siempre bien.

5. Discusión y conclusiones

Lo primero que se puede decir es que, ante un dífícil problema con el que nosotros tendríamos dificultad, es capaz de desarrollar una solución aceptable. Este es un problema que requiere una capacidad de proceso temporal de las señales, la explotación de las características de corporeidad y situación, y la *decisión* de solucionar o no un problema según la situación, para no fallar. En este esquema evolutivo, el fallo implica no sobrevivir, es peor hacerlo mal que no hacerlo. Tampoco hay que olvidar que no sólo se busca que elija una señal, sino que visite las dos en orden, con lo que tiene que cambiar su *estado mental* en función de qué parte de la tarea tiene realizada. Y todo esto con una docena de neuronas.

Es importante señalar que el individuo es un todo, y que su estructura es independiente del tipo de sonido que se use. No se añade ningún tipo de conocimiento para ayudar en la tarea, como ocurre en el reconocimiento de voz, aparte de ser un problema y un planteamiento completamente diferente.

Agradecimientos

Este trabajo está financiado por el proyecto 9/UPV 00003.230-15840/2004.

Queremos también agradecer al grupo *IAS-Research* del departamento de *Lógica y Filosofía de la Ciencia* de la UPV/EHU por sus muy interesantes consejos y ayuda, y a Javier Dolado también por su apoyo y ayuda.

Referencias

1. González-Nalda, P., Cases, B.: Topos: generalized braitenberg vehicles that recognize complex real sounds as landmarks. In Rocha, L.e.a., ed.: Alife X: 10th International Conference on the Simulation and Synthesis of Living Systems, MIT Press (2006, in press)
2. Braitenberg, V.: Vehicles. Experiments in Synthetic Psychology. MIT Press, MA (1984)
3. Brooks, R.: Intelligence without representation. Artificial Intelligence **47** (1991) 139–159
4. Harvey, I., Di Paolo, E., Wood, R., Quinn, M., Tuci, E.A.: Evolutionary robotics: A new scientific tool for studying cognition. Artificial Life **11(1-2)** (2005) 79–98
5. Scutt, T.: The five neuron trick: Using classical conditioning to learn how to seek light. In Cliff, D., Husbands, P., Meyer, J., S.W. Wilson, S., eds.: From Animals to Animats III: SAB'94, MIT Press-Bradford Books, Cambridge, MA. (1994) 364–370
6. Lund, H.H., Webb, B., Hallam, J.: A robot attracted to the cricket species gryllus bimaculatus. In Husbands, P., Harvey, I., eds.: IV European Conference on Artificial Life ECAL97, MIT Press/Bradford Books, MA. (1997) 246–255
7. Yamauchi, B., Beer, R.: Integrating reactive, sequential, and learning behavior using dynamical neural networks. In Cliff, D., Husbands, P., Meyer, J., S.W. Wilson, S., eds.: From Animals to Animats III: on Simulation of Adaptive Behaviour SAB'94, MIT Press-Bradford Books, Cambridge, MA. (1994) 382–391
8. Floreano, D., Mattiussi, C.: Evolution of spiking neural controllers for autonomous vision-based robots. In Gomi, T., ed.: Evolutionary Robotics IV, Berlin, Springer-Verlag. (2001) 3–10
9. Handel, S.: Listening: An Introduction to the Perception of Auditory Events. The MIT Press, Cambridge, MA (1989)
10. Moore, B.C.J.: An Introduction to the Psychology of Hearing. 4th Ed. Academic Press, London (1997)
11. Skinner, B.F.: The behavior of organisms: An experimental analysis. New York: Appleton-Century. (1938)
12. Jakobi, N.: Minimal Simulations for Evolutionary Robotics. PhD thesis. COGS, University of Sussex. (1998)
13. Salomon, R.: The evolution of different neuronal control structures for autonomous agents. Robotics and Autonomous Systems **22** (1997) 199–213
14. Chiel, H., Beer, R.: The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. Trends in Neurosciences **20** (1997) 553–557
15. Maass, W.: Networks of spiking neurons: the third generation of neural network models. Neural Networks **10** (1997) 1659–1671
16. Beer, R., Gallagher, J.C.: Evolving dynamical neural networks for adaptive behavior. Adaptive Behavior **1(1)** (1992) 91–122
17. Maass, W., Bishop, C.M.e.: Pulsed Neural Networks. MIT Press. (1999)

Posprocesamiento morfológico adaptativo basado en algoritmos genéticos y orientado a la detección robusta de humanos

Enrique Carmona, Javier Martínez-Cantos y José Mira

Dpto. de Inteligencia Artificial, ETSI Informática, UNED,
Juan del Rosal 16, 28040, Madrid, Spain.
{ecarmona, jmira}@dia.uned.es
javiermc@info-ab.uclm.es

Resumen. Existen distintas aproximaciones para la detección de objetos móviles basadas en el denominado método de background. En secuencias reales, el gran inconveniente de este método, que comparte con otros métodos de segmentación, es la forma de eliminar el ruido inherente tanto al foreground como al background. Una aproximación muy utilizada para resolver este problema es la aplicación de una secuencia fija de operadores morfológicos pero que, al tenerse que decidirse a priori, no siempre está garantizado el éxito de la restauración del foreground. En este trabajo, se propone un método de posprocesamiento que, para cada frame, ofrece automáticamente una secuencia de operadores morfológicos obtenida a partir de la salida de un algoritmo genético cuya búsqueda está fuertemente sesgada hacia la restauración de siluetas humanas. Finalmente, se propone el uso de este método dentro de un sistema de detección robusta de humanos.

Palabras clave: Segmentación, Morfología, Algoritmos genéticos, Detección de Humanos

1 Introducción

La detección de objetos móviles en secuencias de video es el primer paso relevante en la extracción de información en muchas aplicaciones de visión por computador, incluyendo, por ejemplo, la video-vigilancia, el control de tráfico y el seguimiento de personas. En esta primera etapa de detección, cuanto más fiable sea la forma y posición del objeto que se mueve, más fiable será la identificación del mismo.

En la literatura existen diferentes métodos para la detección de objetos móviles basados, por ejemplo, en métodos estadísticos [1], [2], en la substracción de frames consecutivos [3], en el flujo óptico [4] o en aproximaciones híbridas [5] que combinan algunas de estas técnicas mencionadas. No obstante, por su rapidez y sencillez de implementación, una de las aproximaciones más utilizadas, utilizando cámara fija, se basa en el denominado método de substracción del background y sus múltiples variantes [6], [7], [8], [9], [10], [11], [12]. Básicamente, este método permite detectar regiones en movimiento restando pixel a pixel la imagen actual de una imagen de background tomada como referencia (modelo de background) y creada mediante el

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 249-261, 2006.

promediado de imágenes a lo largo del tiempo durante un periodo de inicialización

Las salidas producidas por los algoritmos de detección mencionados anteriormente, sobre todo si se trabaja con escenas reales, contienen ruido generalmente. Las causas de ruido son debidas principalmente al ruido intrínseco de la propia cámara, a reflejos indeseados, a objetos cuyo color total o parcial coincide con el background y a la existencia de sombras y de cambios (artificiales o naturales) repentinos en la iluminación. El efecto total de estos factores puede ser doble: por un lado, pueden provocar que zonas no pertenecientes a objetos en movimiento se incorporen al foreground (ruido de foreground) y, de otro lado, que determinadas zonas pertenecientes a objetos en movimiento dejen de aparecer en el foreground (ruido de background). En ambos casos, se produce un deterioro en la segmentación de los objetos que es necesario reparar o al menos mejorar si se quiere tener unas garantías mínimas de éxito en las etapas siguientes relacionadas con el *tracking* o con el reconocimiento de objetos.

El recurso empleado en muchos trabajos que se enfrentan con el problema del ruido consiste en la aplicación de una etapa de posprocesamiento basada en la utilización de operadores morfológicos [1], [5], [13], [14], [15]. Por ejemplo, tal y como se indica en la Figura 1, es muy típica la secuencia formada por una fase de dilatación seguida de otra fase de erosión. La primera va encaminada a rellenar todos los huecos que pudieran existir en la silueta de los objetos y/o a conectar distintos fragmentos de un mismo objeto. La segunda es utilizada para invertir la expansión de los límites del objeto producida por la dilatación. Obsérvese que esta decisión tomada anticipadamente puede, en algunos casos, ser crítica. Efectivamente, el gran inconveniente de un esquema fijo de posprocesamiento morfológico radica en su no adaptatividad puesto que, independientemente del mapa de foreground obtenido en un determinado instante del proceso, es necesario haber decidido previamente, en tiempo de compilación, el número de operadores, el orden de ejecución de la secuencia de operadores y el tipo y tamaño del elemento estructurante que utilizará cada uno de ellos.

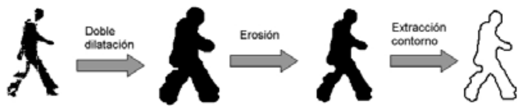


Figura 1. Ejemplo de posprocesamiento morfológico de un objeto de foreground para mejorar su segmentación.

En este trabajo, se propone una nueva aproximación basada en el método de substracción de background combinándolo con técnicas de aprendizaje evolutivas. Básicamente, la idea consiste en hacer evolucionar una población de secuencias de operadores morfológicos mediante un algoritmo genético con el objeto de encontrar la mejor secuencia que permita eliminar el ruido para restaurar el foreground. Para orientar y facilitar la búsqueda, se utiliza una métrica que tiene la propiedad de minimizar su valor cada vez que la secuencia morfológica se aplica a todas aquellas zonas del foreground con siluetas humanas. El resultado final es la construcción de un sistema que abarca simultáneamente la tarea de detección, la de posprocesamiento y la de reconocimiento. Eso sí, las dos últimas, siempre orientadas a la búsqueda de actividad humana.

El resto de este trabajo se organiza de la siguiente manera. En la Sección 2 se aborda la descripción del método de posprocesamiento morfológico evolutivo propuesto, haciendo hincapié en la codificación cromosómica utilizada y en la descripción de la función de adaptación construida. En base al proceso de minimizar esta función, se seleccionará finalmente la secuencia morfológica más apta para restaurar el foreground. La sección 3 describe, mediante un diagrama de bloques la interacción entre la etapa de aprendizaje y el resto de las distintas etapas del sistema propuesto para la detección de humanos. Los resultados experimentales se recogen en la Sección 4 y, finalmente, las conclusiones en la Sección 5.

2 Posprocesamiento morfológico adaptativo basado en algoritmos genéticos

El método de posprocesamiento adaptativo basado en algoritmos genéticos propuesto en este trabajo consiste, básicamente, en buscar en un espacio de secuencias de operadores morfológicos, aquella secuencia óptima que permita mejorar el mapa de foreground en cada instante. Esta búsqueda está fuertemente sesgada (*bias*) debido a que siempre está orientada a la detección de objetos móviles que cumplen unas determinadas características. Concretamente, en el trabajo que aquí nos ocupa, características humanas.

Hay que tener en cuenta que, en general, el modelado de un problema utilizando AG's requiere, en primer lugar, definir la forma de representación de cada individuo (cromosoma), en segundo lugar, construir la función de adaptación necesaria para evaluar la bondad de cada individuo respecto de la solución del problema y, finalmente, elegir los distintos operadores, parámetros y criterios que definen la estructura del AG. Aunque estas tres etapas están fuertemente interrelacionadas, las dos primeras constituyen las etapas más críticas en el proceso de encajar un problema de optimización en el marco de un AG. Así, para la codificación de los individuos, se debe utilizar una representación mínima que sea lo más expresiva posible, de forma tal que sea capaz de representar cualquier solución al problema. En cuanto a la función de adaptación elegida, ésta debe modelar de forma adecuada el problema de optimización que se quiere resolver. El éxito de la solución del problema dependerá en gran medida de ello.

2.1 Codificación cromosómica de la secuencia de operadores

Para cada frame de entrada, tras la fase de detección del foreground, el método aquí propuesto decidirá, en tiempo de ejecución, la secuencia de operadores morfológicos más adecuada a aplicar y orientada a la restauración y detección de todos aquellos objetos que pudieran estar asociados a un humano. En cada instante, esta decisión corresponde a la decodificación del mejor cromosoma perteneciente a la población resultante de la ejecución de un algoritmo genético modelado y parametrizado para tal efecto. Puesto que para definir una secuencia de operadores morfológicos es necesario declarar el tipo de cada operador, su número, el tipo de elemento estructurante

asociado a cada uno de ellos y, finalmente, el orden de aplicación, en esta sección se describe la codificación realizada para definir la estructura cromosómica de cada uno de los individuos que conformarán la población a evolucionar en el AG.

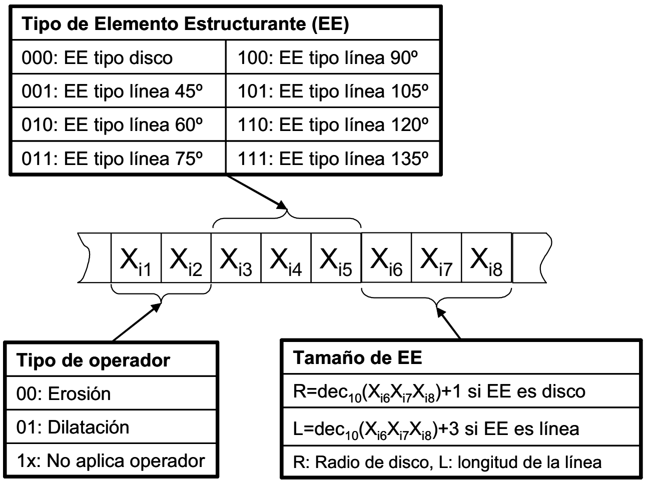


Figura 2. Detalle de la codificación de una secuencia de operadores morfológicos mediante un cromosoma de 64 bits. De izquierda a derecha, cada subcadena de 8 bits del cromosoma contiene la información necesaria para codificar un operador morfológico.

Así, la codificación elegida para cada individuo se hace de acuerdo a la agrupación de 8 subcadenas consecutivas de 8 bits cada una, para formar finalmente una cadena binaria de 64 bits. Por tanto, cada cromosoma codifica una secuencia como máximo de 8 operadores morfológicos. Los dos primeros bits de cada subcadena codifica el tipo de operador morfológico: erosión, dilatación o no aplica operador[1]. De esta forma se consigue que la secuencia de operadores pueda tener un número mínimo de cero operadores y un número máximo de ocho. Los tres siguientes bits codifican el tipo de elemento estructurante: en forma de disco o en forma de línea con diferentes ángulos de inclinación (medidos respecto a la horizontal). La elección del rango de variación de estos ángulos, 45º a 135º, no es arbitraria ya que trata de tener en cuenta el ángulo que puede llegar a formar los brazos o las piernas de un individuo al caminar respecto a su vertical. Finalmente, los tres bits restantes codifican el tamaño del elemento estructurante. En el caso de un elemento estructurante en forma de disco, el radio del disco viene determinado por la decodificación en base decimal de los tres dígitos binarios, sumada en una unidad (no tiene sentido un elemento estructurante de radio 0). Si el elemento estructurante es de tipo línea, la decodificación en base decimal de la terna de bits, aumentada en 3, da el valor de su longitud (no tienen sentido líneas de longitud menor de tres). La Figura 2 muestra esquemáticamente la codificación de un operador morfológico en una subcadena de 8 bits.

[1] Si los dos primeros bits de la subcadena codifican la no aplicación de operador, entonces se ignoran los seis bits restantes de la subcadena.

2.2 Modelado de la función de adaptación

La función de adaptación utilizada por el algoritmo genético va encaminada a evaluar el grado de aproximación entre la silueta de un objeto de foreground restaurado mediante una secuencia de operadores morfológicos y la silueta correspondiente a un humano. Para cuantificar el grado de aproximación entre ambas siluetas, se requiere tanto una métrica adecuada como una base de datos de siluetas humanas patrón con las que realizar la comparación.

La métrica aquí utilizada está basada en la similaridad del contorno de los objetos. Aunque existen numerosos métodos en la literatura para comparar formas [16], [17], aquí nos hemos decantado por la métrica definida en base a la distancia euclídea existente entre la denominada señal de distancia [13] asociada a cada uno de los dos objetos a comparar. Esta elección obedece a la importante propiedad que posee la señal de distancia de ser invariante a las rotaciones, traslaciones y cambio de escala.

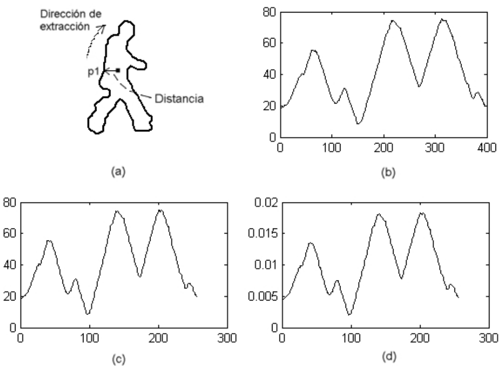


Figura 3. Ejemplo de contorno de un objeto (a) y su correspondiente señal de distancia original (b) escalada (c) y normalizada (d), mostrada en el eje *y* para cada punto del contorno (eje *x*).

Sea S=$\{p_1,p_2,...,p_n\}$ los puntos que determinan el perímetro de un objeto que consta de *n* puntos ordenados, sea p_1 el punto central más a la izquierda de la silueta y el resto obtenidos en la dirección de las manecillas del reloj y sea c_m su centro de gravedad (ver Figura 3). La señal de distancia $SD=\{d_1,d_2,...,d_n\}$, ver ejemplo en Figura 3b, se genera calculando la distancia euclídea entre c_m y cada p_i, es decir,

$$d_i = Dist_{eucl}(c_m, p_i), \quad \forall i \in [1..n]. \tag{1}$$

Para definir una métrica efectiva que permita la comparación de dos formas basada en la similitud de sus señales distancia respectivas, tiene que coincidir la posición relativa, respecto del c.d.g., del primer punto a comparar, el sentido de recorrido del perímetro y el número de puntos en las dos formas a comparar. Cumpliendo estas tres condiciones se consigue que ante dos formas parecidas pero de tamaño diferente, la posición relativa de cada pareja de puntos a comparar sea aproximadamente la misma.

Por tanto, si *N* es el tamaño del vector señal de distancia *SD* y a *C* es una constante usada para fijar el nuevo tamaño del vector a un valor siempre fijo, la señal de distan-

cia escalada, ver ejemplo en Figura 3c, se calcula entonces sub/super muestreando la señal de distancia original *SD* como sigue:

$$\overline{SD}[i] = SD\left[i * \frac{N}{C} \right], \quad \forall i \in [1..C] . \tag{2}$$

En el siguiente paso, la señal de distancia escalada se normaliza para tener integral de área unidad (ver Figura 3c) y garantizar así la invarianza al tamaño. La señal de distancia escalada y normalizada se calcula de acuerdo a la siguiente ecuación:

$$\overline{\overline{SD}}[i] = \frac{\overline{SD}[i]}{\sum_{i=1}^{C} \overline{SD}[i]}, \quad \forall i \in [1..C] . \tag{3}$$

La métrica de clasificación compara la similaridad entre la forma del perímetro de dos objetos, *A* y *B*, calculando la distancia entre sus correspondientes señales de distancia escaladas, SD_A y SD_B. Esta distancia se calcula en [13] como el sumatorio del valor absoluto de las diferencia de las componentes de ambas señales. Aquí se calculará como la distancia euclídea entre las dos señales, es decir:

$$Sim_{AB} = \sqrt{\sum_{i=1}^{i=C} \left(\overline{\overline{SD}}_A[i] - \overline{\overline{SD}}_B[i] \right)^2} . \tag{4}$$

Para completar la definición de la función de adaptación, hay que tener en cuenta que la comparación de la señal de distancia del objeto restaurado por la secuencia de operadores morfológicos que se hace evolucionar, se realiza sobre todas las *M* señales de distancia almacenadas previamente en una base de datos de siluetas humanas. Finalmente, la función de adaptación, f_{adp}, se define como:

$$f_{adp} = \min_{i=1..M} (Sim_{AB[i]}) . \tag{5}$$

Es decir, el AG evoluciona para minimizar la distancia euclídea existente entre la señal de distancia del objeto analizado y la señal de distancia más similar a esta última, existente en la base de datos.

3 Sistema para el procesamiento morfológico adaptativo

El diagrama de bloques de nuestra aproximación para el procesamiento morfológico adaptativo basado en algoritmos genéticos se muestra en la Figura 4. A cada frame de entrada de la secuencia de video, se aplica el método de substracción de background adaptativo para identificar todos aquellos objetos que pertenecen al foreground. La existencia de ruido se manifiesta tanto en el foreground como en el background y, por tanto, se hace necesaria la aplicación de una etapa de posprocesamiento. En nuestro caso, se dividirá en dos fases: la primera encaminada a eliminar el ruido de foreground y, la segunda, a eliminar el ruido de background.

Una de las principales causas que originan ruido de foreground son las sombras que proyectan los objetos y que tienen unos efectos altamente indeseables. Por tanto, se hace necesaria la utilización de algún método [18], [19], [20] para eliminar este tipo de ruido porque, de lo contrario, el fracaso en una posible posterior etapa de reconocimiento de objetos está casi asegurado. En nuestro sistema usamos un esquema de detección y eliminación de sombra basada en el trabajo presentado en [20] que en la práctica demuestra trabajar bien. La primera fase de posprocesado finaliza con la aplicación de un filtro para eliminar todo aquel ruido de foreground asociado a regiones de tamaño muy pequeño y que puede ser debido tanto al propio proceso de segmentación como al resultado de aplicar el filtro de eliminación de sombras.

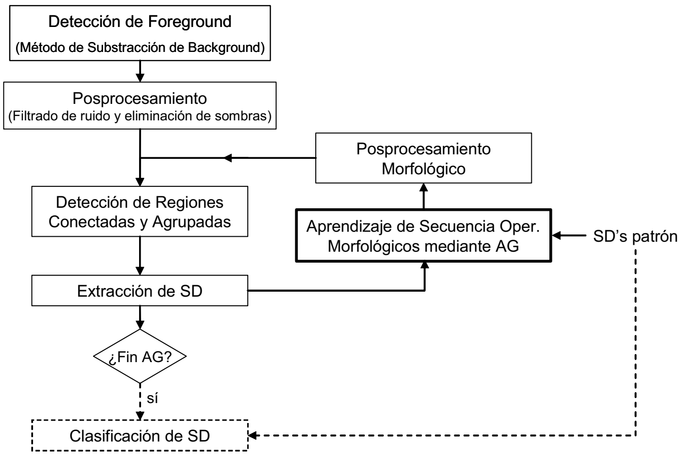


Figura 4. Diagrama de bloques de un sistema de detección robusta de actividad humana en secuencias de vídeo basado en posprocesamiento morfológico adaptativo.

La segunda fase de posprocesado está relacionada con la eliminación de ruido de background expresado por la aparición de huecos en la silueta del objeto, por la disminución de tamaño y/o por la fragmentación del mismo. En este último caso, para determinar las distintas agrupaciones con las que trabajar, se implementa una etapa de detección de regiones conectadas y agrupadas.

En definitiva, la reconstrucción se consigue mediante la aplicación automática del posprocesamiento morfológico adaptativo aquí propuesto y aplicado a cada región de agrupaciones obtenidas en la etapa anterior. En la Figura 4, esto se representa mediante el lazo de realimentación. Así, la misión del módulo de aprendizaje es buscar, desde un punto de vista evolutivo, de forma automática y para cada frame, la mejor secuencia de operadores morfológicos que permita transformar cada agrupación de manchas del foreground, si ello es posible, en una silueta perteneciente a un conjunto de siluetas preestablecido. En nuestro caso, siluetas humanas asociadas a distintas posturas. Una vez que finaliza la búsqueda evolutiva de la secuencia óptima, ésta se aplica.

Es importante resaltar que el método aquí propuesto sólo hace hincapié en la fase de identificación y restauración de objetos móviles en la escena, es decir, no aborda la

tarea de clasificación. Esta es la razón por la que el último módulo de la Figura 4 aparece en línea discontinua. No obstante, con el método propuesto, dicha tarea resultaría relativamente fácil de abordar. Así, se podría medir la similaridad de la señal de distancia del objeto restaurado con las que son de interés y están almacenadas en la base de datos. Para ello, por ejemplo, se podría volver a utilizar la métrica definida en la expresión (5) o, en otro caso, el grado de correlación entre señales distancia. Finalmente, la decisión de pertenencia a la clase se realizaría mediante el uso de un umbral adecuado.

4 Resultados experimentales

Aunque de partida parece más adecuado usar la señal de distancia escalada y normalizada por su invarianza al tamaño, los resultados experimentales demuestran que los mejores resultados se obtienen cuando se trabaja sólo con la señal de distancia escalada sin normalizar. Esto es debido a que se producen restauraciones indeseadas obtenidas como consecuencia de una aplicación excesiva del operador de erosión, que tiende a reducir la silueta del humano hasta una pequeña mancha regular y que, aunque nada tiene que ver con la silueta real, su señal de distancia es muy similar a la que se obtiene para una postura humana con brazos y piernas pegadas al cuerpo. En este sentido, el uso de la señal de distancia escalada sin normalizar implica que la base de datos creada ya sólo es válida para analizar escenas captadas por una cámara que se encuentre a una distancia similar a la que se utilizó para obtener dicha base de datos, es decir, debe cumplirse que los individuos de las secuencias a analizar tengan tamaños similares a los almacenados en la base de datos.

Con la restricción mencionada, para demostrar la validez del método propuesto se utiliza como entrada un conjunto de subsecuencias extraídas de la secuencia denominada *WalkByShop1front.mpg*, perteneciente al almacén[2] público de secuencias de vídeo relacionadas con el denominado proyecto CAVIAR[3]. Así, a modo de ejemplo, la Figura 5 muestra los sucesivos pasos seguidos para procesar uno de los frames perteneciente a lo que hemos denominado *secuencia A*, compuesta de 56 frames. Esta secuencia muestra cómo un hombre, que proyecta sombra, se adentra en el interior de una tienda desplazándose en dirección noreste. Es típico en esta escena el ruido debido a la similaridad de color entre ropa y background. Así, la Figura 5a muestra el ejemplo de frame de entrada seleccionado para esta secuencia y la Figura 5b el foreground resultante de aplicar el método de substracción de background a dicha entrada. Como puede comprobarse, el resultado es altamente ruidoso debido a la existencia de sombras (ruido de foreground). A continuación, la aplicación del filtro de eliminación de sombras produce el resultado mostrado en la Figura 5b. Obsérvese que se ha conseguido reducir el ruido de foreground a costa de producir ruido de background (el objeto disminuye de tamaño y queda fragmentado). Seguidamente, se procede a reconstruir el objeto mediante el método aquí propuesto y cuyo resultado es el que se muestra en la Figura 5d. Finalmente, las Figura 5e-f muestran, respectivamente, la

[2] http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/
[3] http://homepages.inf.ed.ac.uk/rbf/CAVIAR/

señal de distancia de la silueta obtenida tras la restauración y la que de forma automática se selecciona de la base de datos durante el proceso de aprendizaje como señal de distancia más similar. El parecido entre ambas resulta evidente.

Análogamente, en la Figura 6 se muestra el procesado equivalente de uno de los frames pertenecientes, esta vez, a lo que hemos denominado *secuencia B*. La descripción de las distintas subfiguras es totalmente similar a la realizada anteriormente para la Figura 5. Ahora, la *secuencia B* consta de 75 frames y conserva el mismo escenario de background que la anterior. Sin embargo, ahora es una mujer la que se desplaza en dirección este, delante del escaparate de la tienda. En este caso, las causas de ruido son también similares a las de la *secuencia A*: sombras y coincidencia de colores entre objeto y background.

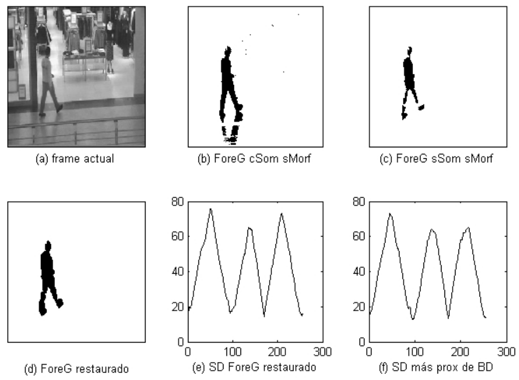


Figura 5. Ejemplo gráfico de la salida de las distintas etapas implicadas en el sistema de procesamiento morfológico adaptativo en la *secuencia B* de vídeo: frame de entrada (a), detección de objetos móviles (b), eliminación de sombras (c), restauración del foreground (d). También se muestra la señal de distancia del humano restaurado (e) y la del humano más similar de la base de datos (f).

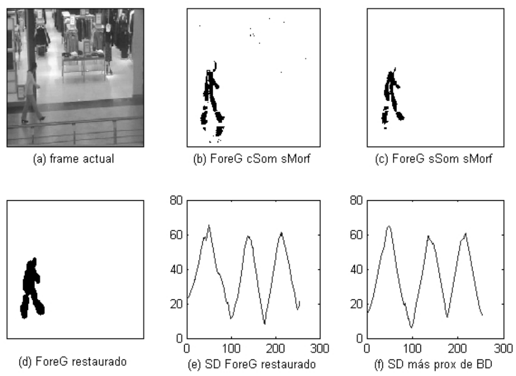


Figura 6. Ejemplo gráfico de la salida de las distintas etapas implicadas en el sistema de procesamiento morfológico adaptativo en la *secuencia A* de vídeo: frame de entrada (a), detección de objetos móviles (b), eliminación de sombras (c), restauración del foreground (d). También se muestra la señal de distancia del humano restaurado (e) y la del humano más similar de la base de datos (f).

Ante la imposibilidad, por motivos de espacio, de mostrar el resultado del procesamiento de todos los frames de las secuencias utilizadas, hemos recurrido a calcular el índice de correlación entre la señal de distancia de la silueta restaurada respecto de la mejor señal de distancia de la base de datos. Los resultados se muestran en la Figura 7 y en ella se puede apreciar, para ambas secuencias, los buenos resultados obtenidos y expresados por el alto grado de correlación conseguido (superior al 0.8 en la mayoría de los frames procesados).

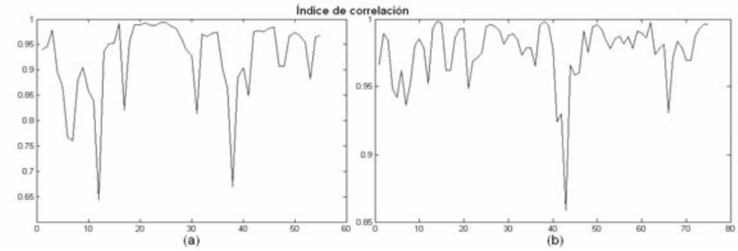


Figura 7. Gráficas que muestran, para cada frame (eje x) de la secuencia, el máximo índice de correlación entre la señal de distancia de la silueta restaurada respecto de la mejor señal de distancia de la base de datos para la *secuencia A* (a) y la *secuencia B* (b).

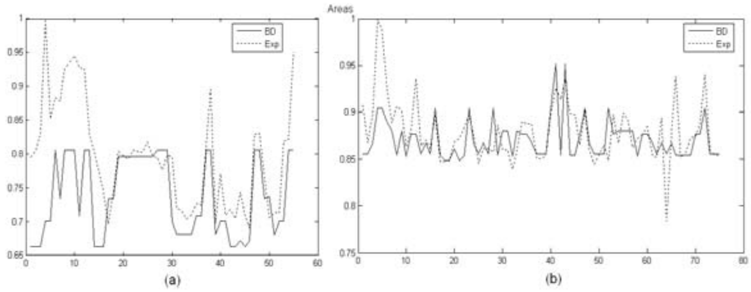


Figura 8. Comparación, para cada frame (eje x) de la secuencia, del área de la silueta restaurada respecto de aquella de la base de datos para la que sus señales distancia respectivas presentan máxima correlación. En cada caso, *secuencia A* (a) y *secuencia B* (b), las áreas se han normalizado respecto al valor máximo obtenido experimentalmente.

Para añadir más información, en la Figura 8 se compara, también para cada frame de cada secuencia, el área de la silueta restaurada (trazo discontinuo) y la correspondiente a la silueta más parecida de la base de datos (trazo sólido), es decir, aquella cuya señal de distancia presenta mayor correlación. Teniendo en cuenta ambos resultados, grado de correlación y área, se observa que en los inicios de ambas secuencias existe un cierto efecto de inercia que impide obtener siluetas restauradas con un valor de área cercano al deseado pero que, sin embargo, permiten ser identificadas como siluetas humanas dado su alto índice de correlación con la base de datos. Es decir, corresponden a siluetas cuya restauración es reconocible pero en las que predomina el efecto de operadores morfológicos de dilatación frente a los de erosión (el área de las siluetas que se obtienen es mayor que el real). La explicación de este fenómeno está motivada por cómo se inicializa la población de operadores morfológicos al analizar

cada frame. En el primer frame de cada secuencia, la población de operadores se escoge de forma aleatoria pero, a partir de aquí, la población final de secuencias de operadores morfológicos obtenida en el frame *i*-ésimo se utilizará como población inicial para procesar el frame *(i+1)*-ésimo. La ventaja de esta política es doble, no sólo ofrece buenos resultados, tal y como lo evidencia el hecho de que las áreas obtenidas se acercan paulatinamente a las áreas reales sino que, además, permite acelerar la convergencia del algoritmo genético en el procesamiento de cada frame.

Finalmente, resaltar que el caso en el que coincide un valor de correlación relativamente bajo con una discrepancia grande de áreas, suele estar asociado a una mala restauración. Este hecho se puede apreciar, principalmente, en el procesado de ciertos frames de la subsecuencia B (Figura 8b).

5 Conclusiones y trabajo futuro

En secuencias de vídeo no captadas en condiciones ideales de laboratorio sino a partir de situaciones reales, el ruido inherente a la propia escena y al originado como resultado de aplicar etapas de procesado a la misma, puede deteriorar la salida resultante de la tarea de segmentación asociada a un sistema de visión de detección de humanos. Aquí se presenta un método novedoso de posprocesamiento morfológico adaptativo basado en algoritmos genéticos y en la idea de señal de distancia. Hay que tener en cuenta que aunque el método presentado está orientado a la restauración de siluetas humanas, el método no pierde generalidad si se quiere aplicar a otro tipo de objetos: simplemente basta utilizar una base de datos de siluetas que sea representativa de la clase de objeto a detectar. Creemos que la aplicabilidad del método radica en facilitar la implementación de una etapa de restauración posterior a cualquier etapa de segmentación y anterior a cualquier etapa de *tracking* o de clasificación. Los prometedores resultados obtenidos avalan la eficacia del método propuesto.

Por otro lado, es conocida la penalización en tiempo asociada al uso de algoritmos genéticos: debería lanzarse un proceso de búsqueda evolutivo por cada frame y por cada región de agrupación de manchas de foreground que haya en la escena. La solución pasa por paralelizar el procedimiento, favorecida además por la naturaleza paralelizable inherente a todo algoritmo genético. Trabajos en este sentido pueden consultarse en [21], [22], [23]. Esta cuestión no ha sido abordada aquí y se propone como una línea de trabajo futura.

Agradecimientos

Esta investigación ha sido soportada por la *CICYT*, a través del proyecto *TIN2004-07661*. También queremos agradecer al *EC Funded CAVIAR project/IST 2001 37540* al hacer público su almacen de secuencias de vídeo, algunas de las cuales han sido utilizadas en este trabajo

Referencias

1. Haritaoglu, I., Harwood, D., Davis, L. S.: "W4: A real time system for detecting and tracking people". In Computer Vision and Pattern Recognition, (1998), 962-967.
2. Stauffer, C., Grimson, W.: "Adaptive background mixture models for realtime tracking". In Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (1999), 246-252.
3. Lipton, A. J., Fujiyoshi, H., Patil, R.S.: "Moving target classification and tracking from real-time video". In Proc. of Workshop Applications of Computer Vision, (1998), 129-136
4. Wang, L., Hu, W., Tan, T.: "Recent developments in human motion analysis". Pattern Recognition, 36 (3), March, (2003), 585-601
5. Collins, R. T. et al.: "A system for video surveillance and monitoring: VSAM final report". Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May (2000).
6. Haritaoglu, I., Harwood, D., Davis, L.S.: "W4: Real-Time Surveillance of People and Their Activities". IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 8, Aug. (2000), 809-830.
7. Amamoto, N., Fujii, A.: "Detecting Obstructions and Tracking Moving Objects by Image Processing Technique". Electronics and Comm. Japan, Part 3, vol. 82, no. 11, (1999), 28-37
8. Stauffer, C., Grimson, W.: "Learning Patterns of Activity Using Real-Time Tracking". IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 8, Aug. (2000), 747-757.
9. McKenna, S.J., Jabri, S., Duric, Z., Rosenfeld, A., Wechsler, H.: "Tracking Groups of People". Computer Vision and Image Understanding, vol. 80, no. 1, Oct. (2000), 42-56.
10. Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.P.: "Pfinder: Real-Time Tracking of the Human Body". IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, July (1997), 780-785
11. Seki, M., Fujiwara, H., Sumi, K.: "A Robust Background Subtraction Method for Changing Background". Proc. IEEE Workshop Applications of Computer Vision, (2000), 207-213
12. Cucchiara, R., Piccardi, M., Prati, A.: "Detecting Moving Objects, Ghost and Shadows in Video Streams". IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 10, Oct., (2003), 1337-1342.
13. Dedeoglu, Y.: "Moving Object Detection, Tracking and Classification for Smart Video Survillance". Ph.D. Thesis, Department of Computer Engineering and The Institute of Engineering and Science of Bilkent University, (2004).
14. Ekinci, M., Gedikli, E.: "Silhouette Based Human Motion Detection and Analysis for Real-Time Automated Video Surveillance". Turk J Elec Engin, vol.13, no. 2, (2005), 199-229
15. Heijden, F.: "Image Based Measurement Systems: Object Recognition and Parameter Estimation". Wiley, January, (1996).
16. Loncaric, S.: "A survey of shape analysis techniques". Pattern Recognition, vol. 31, no. 8, August (1998), 983-1001.
17. Veltkamp, R.C., Hagedoorn, M.: "State-of-the-art in shape matching". Principles of Visual Information Retrieval, Springer, (2001), 87-119.
18. Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sirotti, S.: "Improving Shadow Suppression in Moving Object Detection with HSV Color Information". Proc. IEEE Int'l Conf. Intelligent Transportation Systems, Aug., (2001), 334-339.
19. Prati, A., Cucchiara, R., Mikic, I., Trivedi, M.M.: "Analysis and Detection of Shadows in Video Streams: A Comparative Evaluation". Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, (2001).
20. Horprasert, T., Harwood, D., Davis, L.S.: "A statistical approach for realtime robust background subtraction and shadow detection". In Proc. of IEEE Frame Rate Workshop, Kerkyra, Greece, (1999), 1-19.

21. Bevilacqua, A., “Optimizing parameters of a motion detection system by means of a distributed genetic algorithm”. Image and Vision Computing, vol. 23, (2005), 815-829.
22. Bevilacqua, A., Campanini, R., Lanconelli, N., "A Distributed Genetic Algorithm for Parameters Optimization to Detect Microcalcifications in Digital Mammograms". In Proc. of Third European Workshop on Evolutionary Computation in Image Analysis and Signal Processing (EvoIASP), Como, Italy, LNCS Springer-Verlag, vol. 2037, (2001), 278-287.
23. Cantú-Paz, E., “A Survey of Parallel Genetic Algorithms”. Report No. 97003, citeseer.ist.psu.edu/155991.html.

Mejora paramétrica de la interacción lateral en computación acumulativa

Javier Martínez-Cantos[1], Enrique Carmona[1], Antonio Fernández-Caballero[2] y María T. López[2]

[1] Departamento de Inteligencia Artificial
E.T.S.I. Informática, U.N.E.D, 28040-Madrid, España
javiermc@info-ab.uclm.es,ecarmona@dia.uned.es
[2] Instituto de Investigación en Informática de Albacete (I3A) y
Escuela Politécnica Superior de Albacete
Universidad de Castilla-La Mancha, 02071-Albacete, España
{caballer,mlopez}@info-ab.uclm.es

Resumen El problema de la segmentación de objetos en movimiento en secuencias de vídeo ha sido abordado desde varias aproximaciones. Aumenta un grado la dificultad cuando los objetos monitorizados poseen una apariencia deformable. El método usado en este documento utiliza una red neuronal, explotando la mecánica de la computación acumulativa en conjunción con la interacción lateral recurrente. A pesar de los resultados contrastados en anteriores trabajos, realizamos en este artículo un estudio para mejorar la segmentación sin recurrir a conocimiento de alto nivel. Los módulos propuestos incluyen un filtrado de los objetos según características de tamaño y compacidad y un algoritmo genético capaz de aprender los parámetros que se comportan de un modo mejor.

1. Introducción

El análisis del movimiento visual a partir de imágenes cambiantes en el tiempo es un área importante en visión por computador [2] y en procesamiento de imágenes [10]. Se trata de un único problema con múltiples aplicaciones, al que se destina mucha investigación [9],[10],[4],[1],[13] y que ya ha dado buenos frutos. En particular, los estudios sobre detección de objetos no rígidos están entre los de mayor importancia en análisis del movimiento [4].

Según el enfoque que se utilice en el desarrollo de estos métodos, es posible distinguir entre métodos basados en modelos y métodos guiados por datos. Los primeros, de tipo descendente ("*top-down*"), utilizan conocimiento específico sobre el dominio para construir modelos de aquello que se espera aparezca en la imagen. Luego, se intenta hacer encajar esos modelos con los datos que se presentan en la imagen. El otro tipo de métodos se corresponde con una arquitectura ascendente ("*bottom-up*"). Éstos son apropiados cuando no existe conocimiento sobre qué tipo de objetos pueden aparecer, o bien cuando la

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 262-273, 2006.

diversidad puede ser muy amplia, complicando excesivamente el diseño de un modelo. Estas técnicas operan en tres pasos: preprocesan la imagen para realzar los datos de interés y suprimir el ruido, segmentan los objetos agrupando píxeles pertenecientes a las mismas estructuras en regiones y finalmente interpretan la escena basándose en las características obtenidas.

La interacción lateral en computación acumulativa [5],[6],[7] (de aquí en adelante, ILCA), es un método conducido por datos, capaz de obtener con bastante claridad los objetos deformables presentes en una secuencia de imágenes indefinida, independientemente del tipo de movimiento. La ILCA se implementa como una red neuronal multicapa inspirada en dos modelos: la computación acumulativa local [8] y la interacción lateral recurrente [11]. El método es orientado al píxel y no a regiones, por lo que es más apropiado para ciertos problemas como las oclusiones (ambigüedad del movimiento de los objetos sobre el fondo).

Ahora bien, en un aspecto práctico, tanto las condiciones ambientales, como las distorsiones introducidas por el propio equipo de captación o el tipo de elementos presentes en la escena hacen variar mucho los resultados. La adaptación a estas circunstancias depende de la calibración de los parámetros del sistema. Dicha labor no es automática y requiere un experto que la realice, es decir, un agente externo que interprete la escena a priori y ajuste el sistema para detectar aquello que le interesa. La propuesta presentada en este artículo pretende lograr la autoconfiguración, prescindiendo de conocimiento de alto nivel. Para ello se introducen dos módulos: el primero orientado a mejorar la salida a partir de la incorporación de nuevos parámetros y el segundo dirigido a la realimentación del sistema para aprender los parámetros más adecuados mediante un algoritmo genético al estilo de otros trabajos [3],[14].

2. Breve descripción del método ILCA

Basado en el proceso de visión artificial descrito por Mira y Delgado [12], el sistema se compone de una red neuronal multicapa hacia delante de cuatro capas. Cada píxel en el fotograma de entrada alimenta una neurona en la capa inferior. La capa superior del modelo produce otra imagen de idéntico tamaño, donde se observan un conjunto de siluetas. El método ILCA se ofrece de un modo resumido, ya que puede consultarse en extenso en [7].

2.1. Capa 0: Segmentación por bandas de nivel de gris

Se segmenta la imagen de entrada (en niveles de gris NG) separando en diversas bandas (k) de niveles de gris (BNG) los píxeles que pertenecen a cada una de ellas (ver ecuación 1). Por cada fotograma de la secuencia habrá tantas imágenes como bandas de niveles de gris. El número de bandas de gris n constituye el primero de los parámetros que ofrece la ILCA. Estas bandas tienen el mismo tamaño y no se produce solapamiento entre ellas.

$$BNG_k(x,y,t) = \begin{cases} 1, & \text{si } \frac{NG[x,y,t]}{256} + 1 = k, \forall k \in [0, n-1] \\ -1, & \text{en caso contrario} \end{cases} \tag{1}$$

2.2. Capa 1: Interacción lateral para la computación acumulativa

En esta capa se centra la atención sobre los píxeles que consigan un nivel suficiente de carga de permanencia (CP) calculada a partir de la detección de movimiento a lo largo del tiempo. Denominamos a este método computación acumulativa (ecuación 2). Para ello se recorren todas las bandas, píxel a píxel. Un píxel con carga se identifica como un píxel donde se ha detectado movimiento recientemente. Un píxel donde se acaba de detectar movimiento en el instante actual es cargado al valor de máxima carga o valor de saturación (v_{sat}). Contrariamente, cuando en un píxel no se detecta movimiento, éste se descarga al valor mínimo de carga o valor de descarga (v_{des}). Los píxeles con cierta carga, y en los que se mantiene detección de movimiento, van descargándose gradualmente en un valor v_{dm} de descarga debida al movimiento.

$$CP_k(x,y,t) = \begin{cases} v_{des}, \text{si } BNG_k(x,y,t) = -1 \\ v_{sat}, \text{ si } (BNG_k(x,y,t) = 1) \& (BNG_k(x,y,t-\triangle t) = -1) \\ \text{máx}(CP_k(x,y,t-\triangle t) - v_{dm}, v_{des}), \\ \qquad \text{si } (BNG_k(x,y,t) = 1) \& (BNG_k(x,y,t-\triangle t) = 1) \end{cases} \tag{2}$$

Esta capa dispone de una estructura modular en forma de malla, donde todos los elementos se encuentran interconectados, vertical y horizontalmente, pudiendo comunicarse cada neurona con sus vecinas hasta una distancia de l_1 píxeles a través de canales de entrada y salida. Hablamos de interacción lateral. Un píxel en proceso de descarga puede mantenerse dentro de la silueta del objeto al que pertenece a través de una recarga por vecindad (v_{rv}), pues los píxeles con máxima carga actúan como iniciadores de una interacción lateral, que transcurre a través de todos los píxeles cuya carga no sea absoluta (ni v_{sat}, ni v_{des}). Por eso, se dice que se comportan como estructuras transparentes. Del mismo modo, los píxeles con carga mínima paran el avance: son estructuras opacas. La ecuaciones 3 y 4 describen este comportamiento.

$$CP_k(x,y,t) = min(CP_k(x,y,t) + \epsilon \cdot v_{rv}, v_{sat}) \tag{3}$$

donde

$$\epsilon = \begin{cases} 1, \text{ si } \exists (i \le l_1) | \forall (1 \le j \le i) \\ \quad ((CP_k(x+i,y,t)) = v_{sat} \bigcap (CP_k(x+j,y,t)) \neq v_{des} \bigcup \\ \quad (CP_k(x-i,y,t)) = v_{sat} \bigcap (CP_k(x-j,y,t)) \neq v_{des} \bigcup \\ \quad (CP_k(x,y+i,t)) = v_{sat} \bigcap (CP_k(x,y+j,t)) \neq v_{des} \bigcup \\ \quad (CP_k(x,y-i,t)) = v_{sat} \bigcap (CP_k(x,y-j,t)) \neq v_{des}) \\ \\ 0, \text{ en caso contrario} \end{cases} \tag{4}$$

Por último, se aplica un valor umbral denominado valor mínimo de mancha por banda de nivel de gris (θ_{per}). Con todo ello, se obtiene el valor de permanencia final.

2.3. Capa 2: Interacción lateral para la obtención de elementos de siluetas

Los valores de permanencia calculados por la capa 1 son ofrecidos a esta capa (ahora las cargas de permanencia pasan a denominarse C), donde de nuevo se presenta una estructura modular en forma de malla. En esta etapa, la carga es repartida entre todos los píxeles (en una distancia máxima l_2) que forman una silueta, entendiendo como tal al conjunto de los píxeles vecinos, dentro de la misma banda, que tengan carga no nula. La interacción lateral se encargará de delimitar esos repartos y de repartir uniformemente la carga dentro de cada mancha. Así se definen las siluetas de los objetos, se diluye el movimiento del fondo y se obtiene cierta aproximación a la clasificación de los objetos basándose en el color de las manchas (ver ecuación 5). En esta capa también existe un umbral final que restringe la salida a la siguiente capa, a saber, el valor mínimo de mancha para la fusión de objetos (θ_{car}).

$$C_k(x,y,t) = \frac{C_k(x,y,t) + \sum_{i=-l_2}^{l_2} \sum_{i=-l_2}^{l_2} \delta_{x+i,y+j} \cdot C_k(x+i,y+j,t)}{1 + \sum_{i=-l_2}^{l_2} \delta_{x+i,y+j}}, \quad (5)$$

$$\forall (i,j) \neq (0,0)$$

donde

$$\delta_{\alpha,\beta} = \begin{cases} 1, & \text{si } C_k(\alpha,\beta,t) > v_{des} \\ 0, & \text{en caso contrario} \end{cases} \quad (6)$$

2.4. Capa 3: Interacción lateral para la fusión de objetos en movimiento

Por último, se reúnen de nuevo todas las subcapas para generar la imagen final S, según muestra la fórmula 7.

$$S(x,y,t) = max(C_k(x,y,t)), \forall k \in [0,255] \quad (7)$$

Se procede, aplicando a cada píxel de la imagen final el valor máximo entre los correspondientes a las mismas coordenadas, en cada subcapa anterior. Posteriormente, se realiza la media de cada punto con los vecinos del entorno (de nuevo, mediante la interacción lateral hasta una distancia de l_3 píxeles)(ecuaciones 8 y 9). Finalmente, se aplica el último de los umbrales (θ_{obj}), llamado valor mínimo de detección de siluetas.

$$S(x,y,t) = \frac{S(x,y,t) + \sum_{i=-l_3}^{l_3} \sum_{i=-l_3}^{l_3} \delta_{x+i,y+j} \cdot S(x+i,y+j,t)}{1 + \sum_{i=-l_3}^{l_3} \delta_{x+i,y+j}}, \quad (8)$$

$$\forall (i,j) \neq (0,0)$$

donde

$$\delta_{\alpha,\beta} = \begin{cases} 1, & \text{si } S(\alpha,\beta,t) > v_{des} \\ 0, & \text{en caso contrario} \end{cases} \quad (9)$$

3. Mejora paramétrica de la ILCA

En esta sección se presenta un marco de trabajo que incluye el método ILCA (en sus cuatro capas) y añade unos módulos externos a la misma para la mejora paramétrica del método. Dicha mejora tendrá en cuenta la escena específica tratada. El marco completo del sistema de segmentación se muestra en la figura 1.

Como se ha visto, la ILCA produce conjuntos de siluetas para cada fotograma de la secuencia que procesa. El módulo "discriminación de objetos"filtra las siluetas, según criterios del usuario, y, dependiendo de la escena específica, para obtener sólo los objetos de interés en cada una de las imágenes. El módulo "refinamiento de parámetros"manipula los parámetros de la ILCA basándose en el número de objetos detectados frente a los realmente de interés (dato indicado por el usuario). La composición de los parámetros se realiza aplicando un algoritmo genético.

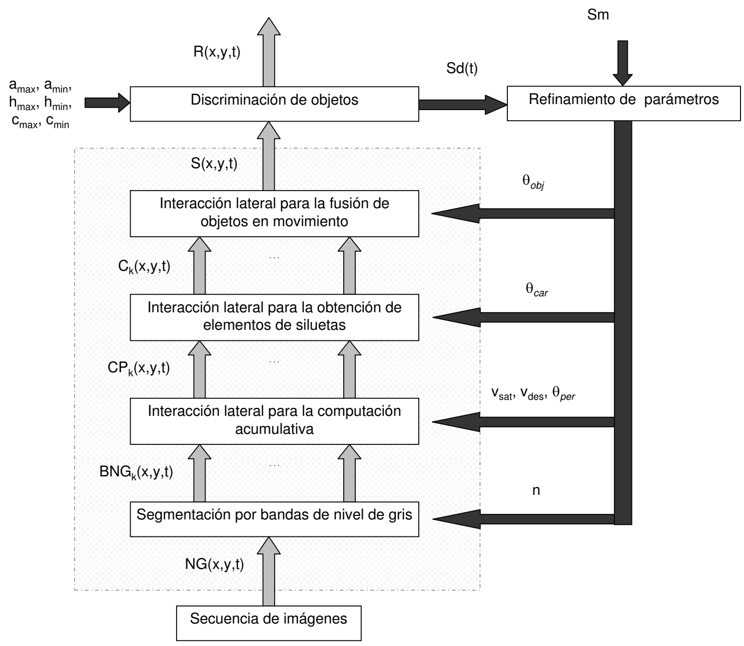


Figura 1. Marco de trabajo para la solución propuesta

3.1. Discriminación de objetos

El conjunto de siluetas resultante del proceso de ILCA es filtrado por medio de los criterios de “tamañoz “compacidad”. Cada escena específica monitorizada marca en qué márgenes se encuentran los objetos de interés (en píxeles): anchura máxima (a_{max}), anchura mínima (a_{min}), altura máxima (h_{max}) y altura mínima (h_{min}). Otro factor que puede actuar en conjunción es el porcentaje que ocupan los objetos dentro de la caja (*bounding box*) que los rodea: hablamos de la compacidad máxima (c_{max}) y de la compacidad mínima (c_{min}). La escena resultante de todo el proceso es almacenada junto con la contabilización, en cada fotograma, del número de objetos detectados (Sd). Este módulo reduce la rigurosidad con que debe ser configurada la ILCA, pues puede filtrar algunos objetos no buscados o ruido.

3.2. Refinamiento de parámetros

La estructura de la ILCA corresponde a una red neuronal y, por tanto, conlleva un sistema de aprendizaje de la misma. Tratamos de dotar de algún mecanismo que realimente el ciclo, desde la capa inferior, permitiendo modificar los parámetros de configuración. Se utiliza un algoritmo genético, por su idoneidad en la búsqueda de soluciones en problemas de optimización donde el espacio de búsqueda es tan amplio que no permite un recorrido exhaustivo. El usuario debe orientar al algoritmo genético, indicando cuántos objetos en movimiento hay en la imagen, o más correctamente, cuántos le interesan.

El algoritmo genético asistirá al sistema en la búsqueda no supervisada de parámetros adecuados según el usuario establezca: tamaño (T) de la población (conjunto de soluciones), puntos de recombinación crossover en la reproducción, probabilidad de mutación, número de generaciones (N) y número de objetos esperados (Sm). Así el algoritmo genético procesará la secuencia original $T+N*(T/2)$ veces, con los parámetros que vayan determinando los individuos de la población. La ILCA es configurada en base a seis parámetros distintos (vistos con anterioridad) que formarán los cromosomas de los individuos: número de bandas de nivel de gris (n), descarga debida al movimiento (v_{dm}), recarga debida a la vecindad (v_{rv}), valor mínimo de mancha por banda de nivel de gris (θ_{per}), valor mínimo de mancha para la fusión de objetos (θ_{car}) y valor mínimo de detección de siluetas (θ_{obj}).

Según la regla de los bloques de construcción, es recomendable situar en genes consecutivos los parámetros relacionados entre sí. El número de bandas de nivel de gris (n), aparentemente, no tiene relación con el resto. Descarga debida al movimiento (v_{dm}) y recarga por vecindad (v_{rv}) sí que parecen mucho más afines y consecuentemente deberían ir uno junto al otro. Los tres últimos parámetros citados son umbrales (θ_{per}, θ_{car}, θ_{obj}). Aunque en principio parecen no guardar una relación directa, puede ser interesante reunirlos, pues todos ellos representan umbrales que determinan qué valores pasan a la capa siguiente. En consecuencia, la población se codificará en el orden (n, v_{dm}, v_{rv}, θ_{per}, θ_{car}, θ_{obj}). La función de evaluación seleccionada (E) pretende minimizar el error producido

por el número de objetos detectados en cada instante ($Sd(t)$) respecto al número de objetos esperados por el usuario (Sm), para una secuencia de k fotogramas, conforme se observa en la ecuación 10:

$$E = \sum_{t=0}^{k-1} \frac{|Sd(t) - Sm|}{|t - \frac{k}{2}| + 1} \tag{10}$$

Así pues, esta función de adaptación o fitness considera en cada fotograma la diferencia entre el número de objetos detectados y los esperados, dotando de mayor importancia a los fotogramas centrales de la secuencia, ya que habitualmente al comienzo y al final de las secuencias no se visualizan todos los objetos o es difícil detectarlos, además de que al principio, el algoritmo requiere de algunos fotogramas para converger.

4. Datos y resultados

Por último, se realiza un análisis de los resultados producidos según se manipulen los módulos de "discriminación de objetosz de "refinamiento de parámetros". Para ello se utiliza una secuencia formada por 49 fotogramas, en 256 niveles de escala de gris y con dimensiones 128x128 píxeles. Se trata de una escena sencilla, donde aparece un humano desplazándose a lo largo de una habitación (ver figura 2).

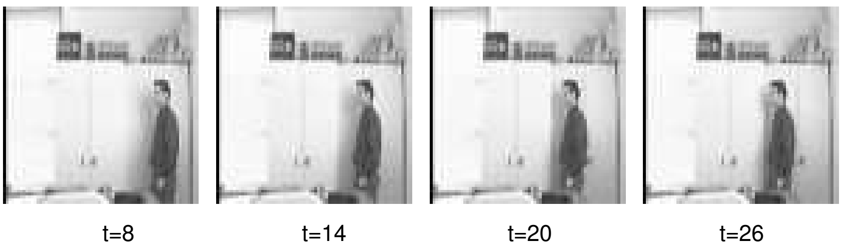


Figura 2. Algunas tramas de la secuencia de entrada

Como ya se ha mencionado, la ILCA ofrece resultados aceptables cuando los parámetros están bien configurados [7]. El número de parámetros del método ILCA es elevado y su ajuste es complicado. Por ello, históricamente se suele acudir a un mismo conjunto cuyo comportamiento es satisfactorio en muchas escenas. Típicamente se vienen utilizando 8 bandas de nivel de gris, 63 como valor de descarga debida al movimiento, 31 como valor de recarga debida a la vecindad y 150 para cada uno de los umbrales, es decir, valor mínimo de mancha por banda de nivel de gris, valor mínimo de mancha para la fusión de objetos y valor mínimo de detección de siluetas, respectivamente. Según la descripción ofrecida acerca del "refinamiento de parámetros", el cromosoma típico que configura la ILCA es

(8, 63, 31, 150, 150, 150). En la figura 3 aparece el resultado del procesamiento de la secuencia ejemplo con este cromosoma.

Efectivamente se obtiene la silueta del objeto en movimiento; sin embargo alrededor de ella aparece mucha información extraña. La inclusión de ruido en cantidad es uno de los principales problemas que se derivan de la aplicación de unos parámetros poco efectivos. A pesar de no ser percibible visualmente, en cada fotograma existen alrededor de 500 objetos detectados. Esto dificulta el tratamiento de la información por capas de software de más alto nivel que utilicen la ILCA como base. Además, la silueta es confusa, pues unido a los contornos del objeto que se desplaza, existen otros contornos que pertenecen a elementos del fondo de la imagen o incluso a su propia sombra.

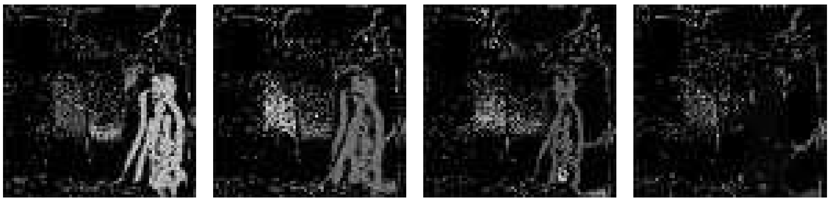


Figura 3. Procesado con parámetros típicos ($fitness = 2754'41$)

Los resultados se acompañan de la medida de fitness para poder comparar los diversos métodos de un modo más riguroso. En este caso la medida es muy alta, considerando que un valor cero representa que se detectaron únicamente los objetos indicados: en el ejemplo buscamos sólo un objeto ($Sm = 1$), es decir, al humano.

4.1. Resultados tras la "discriminación de objetos"

Hemos indicado anteriormente que el filtrado de objetos puede realizarse según dos criterios: compacidad y tamaño. En el primero de los casos, el usuario se encarga de establecer la proporción de espacio que un objeto puede ocupar dentro de la caja que lo delimita. La figura 4a muestra el resultado de procesar la secuencia con el mismo cromosoma pero con una limitación de compacidad máxima del 95 %.

Se observa con claridad cómo ha desaparecido de los fotogramas mucho ruido. Aunque no consta en los fotogramas presentados, este factor también tiene efecto beneficioso sobre el proceso de convergencia de la secuencia, pues evita la detección de movimiento en el primer fotograma debido a la carga inicial de todos los píxeles. La mejoría es visible y así lo indica la reducción del fitness en más de siete veces su valor anterior. Por su parte, el establecimiento del valor menor de compacidad también es importante. Sobre la secuencia procesada con

el cromosoma típico aplicamos ahora un límite mínimo de compacidad del 40 %, sin imponer un máximo.

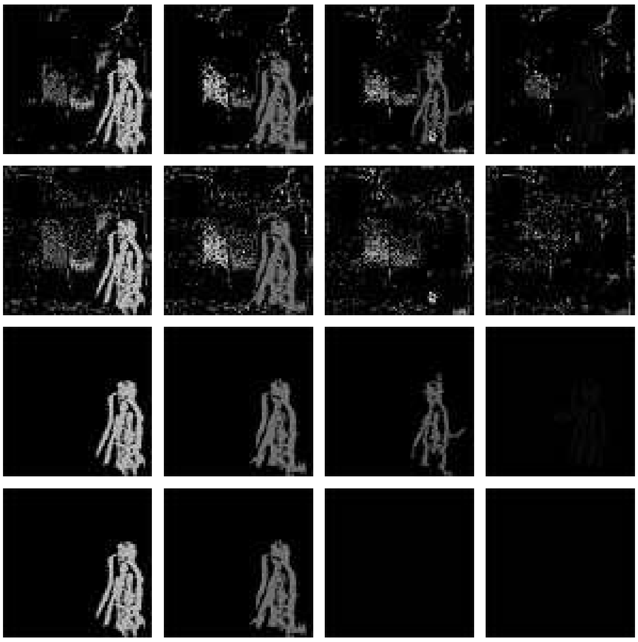

(a)
(b)
(c)
(d)

Figura 4. Mejora por compacidad y tamaño. (a) Compacidad máxima 95 % ($fitness = 377'45$). (b) Compacidad mínima 40 % ($fitness = 2721'21$). (c) Altura 60-100 y anchura 25-90 ($fitness = 0'167$). (d) . Compacidad 40-95 %, altura 60-100 y anchura 25-90 ($fitness = 4'854$)

En la figura 4b aparece el resultado. En este punto queda de manifiesto la relevancia de una configuración acertada en los parámetros de la ILCA. Anteriormente se visualizaba en todos los fotogramas el objeto en movimiento, además de otros tantos no deseados. Ahora, por la acción de la compacidad mínima, la persona queda filtrada en algunas imágenes. Esto se debe a que los objetos no son detectados con exactitud y con frecuencia son encerrados en grandes cajas junto con otros elementos extraños, formando una única silueta. Por este motivo, sucede que algunas cajas contienen al humano y éste ni tan siquiera ocupa el 40 % de ese espacio.

Aunque el resultado no es demasiado bueno, se opta por mantener este parámetro de compacidad a dicha cantidad pues la pretensión es obtener siluetas que comprendan mayoritariamente al objeto que representan. El otro tipo de filtrado restringe el tamaño de estas cajas. Según la aplicación deberán medirse los objetos monitorizados, en píxeles, para especificar a partir de qué tamaño un

objeto es interesante. Así será posible evitar la interferencia de pequeños objetos y en general de ruido. De forma análoga, será conveniente establecer el tamaño máximo. La figura 4c presenta el resultado para cajas de anchuras entre 25 y 90 píxeles, y alturas entre 60 y 100 píxeles, sin restricciones de compacidad.

Ahora, el resultado ha mejorado ostensiblemente. El ruido ha desaparecido por completo y el fitness presenta una medida muy buena. También desaparece el efecto de la convergencia. Sin embargo, sólo se ha limpiado el resultado del procesado original. Las siluetas representan claramente la posición del objeto en movimiento, mostrando incluso algunos de sus contornos principales, pero continúan existiendo bandas unidas al objeto que no forman parte de él. Ha sido posible extraer aquellos objetos de interés de entre los generados, pero sigue siendo necesario un modo de mejorarlos. En adelante aplicaremos ambos métodos de discriminación de un modo conjunto, aunque dificulten a priori la detección de objetos, tal como se observa en la figura 4d y sobre todo en su fitness: en algunos fotogramas el objeto de interés es filtrado por efecto de la compacidad mínima.

4.2. Resultados tras el "Refinamiento de parámetros"

El algoritmo genético es el medio idóneo para obtener buenas configuraciones para la ILCA. Limitando la compacidad entre 40 y 95 %, la altura entre 60 y 100 píxeles, y la anchura entre 25 y 90 píxeles, se ha ejecutado en varias ocasiones este módulo. Generalmente, las poblaciones utilizadas fueron de 16 individuos o cromosomas, con crossover de 3 puntos y probabilidad de mutación del 8 % por gen. El algoritmo ha sido ejecutado durante 14 generaciones, calculando un total de 128 individuos cada vez. En algunas de estas ejecuciones se introdujeron en la población inicial determinados cromosomas para sesgar la evolución, por ejemplo, el cromosoma típico (8, 63, 31, 150, 150, 150). A excepción de esos cromosomas, la población inicial es generada al azar: todos los parámetros varían entre 0 y 255, menos el número de bandas por nivel de gris (n) que sólo toma valores 2, 4, 8 o 16 por motivos de eficiencia.

En la figura 5 aparecen algunos fotogramas característicos de los cromosomas obtenidos de este modo. A pesar de no haber producido muchas generaciones, el algoritmo genético ofrece algunos resultados interesantes. Se han alcanzado medidas de fitness relativamente bajas y, exceptuando algunos fotogramas donde se pierde el objeto debido a las restricciones del módulo de discriminación, los resultados son satisfactorios. Dependiendo del cromosoma utilizado, la silueta se presenta más o menos definida, pero siempre suavizada y sólida. Debe observarse que en algunos fotogramas donde antes se perdía el objeto (figura 4d, $t = 20$ y $t = 26$), ahora no ocurre, pues la silueta producida es más perfecta y no presenta contornos del fondo unidos a ella.

5. Conclusiones

La monitorización con cámaras fijas se caracteriza por vigilar espacios cuyas condiciones ambientales están controladas y son poco variables. Bajo estas

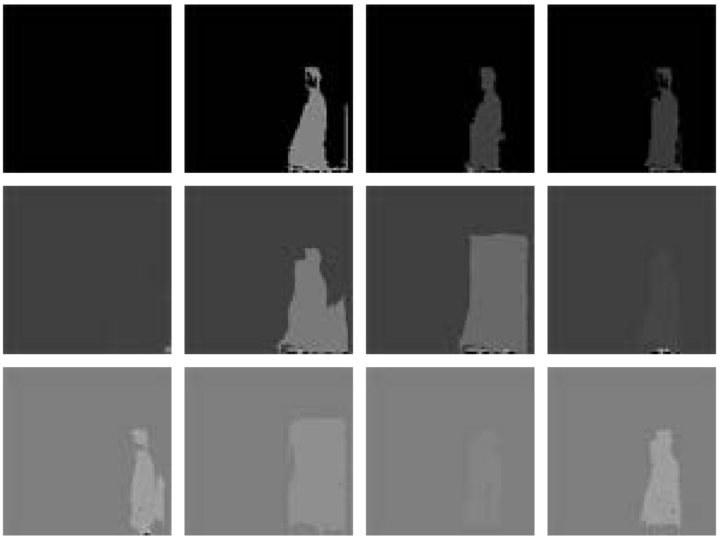


Figura 5. Resultados con distintos cromosomas. (a) (4, 63, 106, 99, 150, 36) ($fitness = 2'753$). (b) (8, 30, 46, 105, 173, 31) ($fitness = 2'197$). (c) (8, 102, 200, 37, 210, 14) ($fitness = 0'8781$)

condiciones, la ILCA dispone de capacidad suficiente para adaptarse a las pequeñas variaciones que puedan producirse en tal situación, siempre y cuando se parta de una buena configuración.

Entonces, el problema es encontrar un conjunto de parámetros adecuado para el escenario elegido. El modelo propuesto configura automáticamente el sistema tomando una secuencia captada en el lugar donde vaya a implantarse. Además, añade un mecanismo para relajar la rigurosidad del proceso, pues se efectúan post-procesados para suprimir objetos indeseados. El algoritmo genético muestra un panorama alentador, pues con pequeñas pruebas genera resultados esperanzadores. Parece interesante continuar en el mismo camino, probando con otras configuraciones de ambos módulos, para conocer un tanto mejor la composición de los cromosomas. También conviene intentar conducir las ejecuciones, combinando varios cromosomas de buen comportamiento y obtener poblaciones más refinadas.

Agradecimientos

Este trabajo ha sido parcialmente financiado por los proyectos CICYT TIN2004-07661-C02-01 y TIN2004-07661-C02-02.

Referencias

1. Aggarwal, J.K., Nandhakumar, N.: On the computation of motion from sequences of images - A review. Proceedings of the IEEE (1988) 917–935
2. Bathe, K.: Finite Element Procedures in Engineering. Prentice-Hall (1982)
3. Chiu, P., Girgensohn, A., Polak, W., Rieffel, E.G., Wilcox, L., Bennett, F.H. III: A genetic segmentation algorithm for image data streams and video. Proceedings of the Genetic and Evolutionary Computation Conference (2000) 666–673
4. Faugeras, O.D., Lustman, F., Toscani, G.: Motion and structure from motion from point and line matches. Proceedings of the 1st International Conference on Computer Vision (1987) 25–34
5. Fernández-Caballero, A., Mira, J., Fernández, M.A., López, M.T.: Segmentation from motion of non-rigid objects by neuronal lateral interaction. Pattern Recognition Letters **22**:14 (2001) 1517–1524
6. Fernández-Caballero, A., Mira, J., Delgado, A.E., Fernández, M.A.: Lateral interaction in accumulative computation - A model for motion detection. Neurocomputing **50**C (2003) 341–364
7. Fernández-Caballero, A., Fernández, M.A., Mira, J., Delgado, A.E.: Spatio-temporal shape building from image sequences using lateral interaction in accumulative computation. Pattern Recognition **36**:5 (2003) 1131–1142
8. Fernández, M.A., Mira, J.: Permanence memory - A system for real time motion analysis in image sequences. Proceedings of the IAPR Workshop on Machine Vision Applications (1992) 249–252
9. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artificial Intelligence **17** (1981) 185–203
10. Jain, A.K.: Fundamentals of Digital Image Processing. Prentice-Hall (1989)
11. Mira, J., Delgado, A.E., Manjarrés, A., Ros, S., Alvarez, J.R.: Cooperative processes at the symbolic level in cerebral dynamics - Reliability and fault tolerance. Brain Processes Theories and Models, MIT Press, Cambridge, MA (1996) 244–255
12. Mira, J., Delgado, A.E., Boticario, J.G., Díez, F.J.: Aspectos básicos de la inteligencia artificial. Editorial Sanz y Torres, S. L. Madrid (1995)
13. Mitiche, A., Bouthemy, P.: Computation and analysis of image motion - A synopsis of current problems and methods. International Journal of Computer Vision **19**:1 (1996) 29–55
14. Ramos, V., Muge, F.: Image colour segmentation by genetic algorithms. Proceedings of the 11th Portuguese Conference on Pattern Recognition (2000) 125–129

Aprendizaje de reglas difusas ponderadas mediante algoritmos de estimación de distribuciones

Luis delaOssa, José A. Gámez y José M. Puerta

Grupo de investigación en Sistemas Inteligentes y Minería de Datos / $i^3\mathcal{A}$
Departamento de Sistemas Informáticos - Universidad de Castilla-La Mancha
Campus Universitario s/n, 02071, Albacete (España)
[ldelaossa|jgamez|jpuerta]@info-ab.uclm.es

Resumen En los Algoritmos de Estimación de Distribuciones, la evolución se lleva a cabo mediante el aprendizaje de una distribución de probabilidad a partir de las mejores soluciones en una población, y la generación de nuevas soluciones mediante el muestreo de esta. En este trabajo se estudia la aplicación de estos algoritmos al aprendizaje de sistemas basados en reglas difusas ponderadas mediante la metodología WCOR. Para ello, proponemos el uso de dos modelos probabilísticos diferentes, uno que no asume ningún tipo de dependencia entre los consequentes de las reglas y sus variables y otro cuya estructura es fijada a partir de estas dependencias conocidas a priori.

1. Introducción

Un *sistema difuso genético* (SDG) [1] o *sistema difuso evolutivo* (SDE), en un sentido más amplio, es un sistema difuso inducido a partir de datos mediante el uso de un algoritmo genético (evolutivo). Aunque existen diferentes tipos, los *sistemas basados en reglas difusas* (SBRDs) son los que han recibido mayor atención de la comunidad de SDEs. En este caso, los algoritmos evolutivos son usados para aprender o ajustar diferentes componentes de las reglas.

En este trabajo, nos centraremos en un tipo concreto de SBRD: el que usa reglas difusas *lingüísticas* o *descriptivas* [2]. Los SBRD lingüísticas (SBRDLs) son especialmente atractivos por que permiten lograr el doble objetivo de ser predictivos y a la vez plenamente interpretables por los expertos humanos.

La capacidad de predicción de los SBRDLs puede ser, en general, mejorada mediante el uso de pesos. Un peso es un número real, $w \in [0, 1]$, que se asocia a cada regla y puede ser entendido como su grado de importancia. Esta técnica, además, no perjudica la interpretabilidad del sistema de manera significativa. La metodología WCOR (*Weighted Cooperative Rules* o Reglas Ponderadas Cooperativas)[3,4], se centra en el aprendizaje de las reglas y sus pesos, pero no lleva a cabo ningún ajuste de la forma de los conjuntos difusos asociados a las etiquetas lingüísticas.

El objetivo principal de este estudio es analizar las posibilidades que ofrece el uso de *Algoritmos de Estimación de Distribuciones* (AEDs) [5] como motor de búsqueda en wCOR. Para ello, se han elaborado dos propuestas que pueden ser vistas como la

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 274-285, 2006.

correspondencia a los algoritmos descritos en [4], ya que evolucionan simultáneamente las reglas y sus pesos.

El artículo comienza con una descripción breve de los SBRDLs (Sección 2) y como aprenderlos mediante el uso de la metodología WCOR (Sección 3). Después, en la sección 4 se describe el AED canónico y los algoritmos usados en este estudio. La sección 5 contiene los algoritmos propuestos para tratar el problema del aprendizaje de RDLs y en la Sección 6 se evalúan con una serie de conjuntos de datos. Finalmente, en la Sección 7 se presentan las conclusiones y el trabajo futuro.

2. Sistemas basados en reglas difusas lingüísticas

Las Reglas Difusas (RDs) [6] se basan en la *Teoría de Conjuntos Difusos* [7] y se fundamentan en el uso de predicados difusos, X es A, donde X es una variable del dominio del problema y A es un conjunto difuso. La estructura típica de una regla difusa es

$$\text{If } X_1 \text{ es } v_1^{j_1} \& \ldots \& X_n \text{ es } v_n^{j_n} \text{ entonces } Y \text{ es } v_y^{j_l} \text{ con peso } w \in [0,1] \quad (1)$$

donde $\{X_1, \ldots, X_n, Y\}$ son las variables de dominio del problema y $\{v_1^{j_1}, \ldots, v_n^{j_n}, v_y^{j_l}\}$ son conjuntos difusos definidos sobre el dominio de las variables correspondientes y cada regla tiene asociado un peso w o grado de importancia.

Como se mencionó en la Sección 1, nos centramos en las llamadas reglas difusas *lingüísticas* o tipo *Mamdani* [2]. La base de conocimiento del SBRDL se compone de dos elementos claramente diferenciados:

• Una *base de datos lingüística* que contiene la definición de las variables. Esto es, el dominio de cada variable de entrada/salida es partido/cubierto por un número fijo de conjuntos difusos, y cada uno tiene asociado una variable lingüistica. Por ejemplo, el dominio de la variable *edad*, puede ser cubierto por el conjunto de etiquetas lingüísticas: {bebe, niño, adolescente, adulto, anciano}. Mediante la asociación de un conjunto difuso a cada etiqueta lingüística se tiene una *variable lingüística* [8].

• Una *base de reglas* definida sobre las variables lingüísticas. En el *modelado lingüístico* de un sistema, sólamente las etiquetas lingüísticas de una variable pueden aparecer en los predicados difusos de las reglas. Esto es, el conjunto difuso $v_1^{j_1}$ usado para la variable X_1 en la equación 1 no puede ser elegido con total libertad, sino de entre el conjunto de etiquetas lingüísticas usadas en la definición de la variable lingüística X_1.

Por esta restricción, los SBRDLs tienen habitualmente una precisión menor que otros tipos de sistemas basados en reglas difusas pero, por otra parte, tienen una alta interpretabilidad.

Con respecto a la inferencia, pueden distinguirse las siguientes componentes:

• *Interfaz de fuzzificación / defuzzificación*. El primer paso es la transformación de los valores numéricos en conjuntos difusos, llevada a cabo mediante la obtención de un conjunto difuso puntual $\hat{r}$ dado un número r, es decir, un conjunto difuso tal que r tiene un grado de pertenencia de 1 ($\mu_{\hat{r}}(r) = 1{,}0$) y cada punto tal $s \neq r$ tiene un grado de pertenencia 0. Por otra parte, la interfaz de defuzzificación toma un conjunto difuso y produce una salida numérica mediante el uso (en este caso) del centro de gravedad del conjunto difuso dado.

• *Motor de inferencia*. Dada una entrada $\mathbf{x} = \langle x_1, \ldots, x_n \rangle$ cualquier regla (ver eq. 1) tal que $\forall_{i=1\ldots n}\ \mu_{v_i^{j_i}}(x_i) > 0$ se dispara. Como los conjuntos difusos que definen las variables difusas generalmente se solapan, una entrada generalmente dispara varias reglas. Cuando una regla se dispara, se obtiene un conjunto difuso para la variable objetivo (Y). En este trabajo, el conjunto v'_y se obtiene (mediante operadores clásicos) como:

$$\mu_{v'_y}(r) = \begin{cases} \mu_{v_y^{j_l}}(r) & \text{if } \mu_{v_y^{j_l}}(r) < m \\ m & \text{if } \mu_{v_y^{j_l}}(r) \geq m \end{cases}$$

siendo m el grado de emparejamiento de $\mathbf{x}$ con la regla: $m = min_{i=1..n}\mu_{v_i^{j_i}}(x_i)$.

Si k reglas son disparadas por una entrada dada $\mathbf{x}$ y $v_y^1, \ldots, v_y^k$ son los conjuntos difusos obtenidos, entonces tenemos que combinarlos en una única salida. En este sentido, usamos la aproximación FITA (*First Integrate Then Aggregate*) ponderada, que primero defuzzifica $v_y^1, \ldots, v_y^k$ a sus valores numéricos correspondientes $r_y^1, \ldots, r_y^k$ y entonces los agrega a un valor numérico único mediante el uso de la media ponderada teniendo en cuenta el peso (w_i) asociado a cada regla:

$$\hat{y} = \frac{\sum_{i=1}^{k} r_y^i \cdot m_i \cdot w_i}{\sum_{i=1}^{k} m_i \cdot w_i},$$

siendo m_i el grado de emparejamiento de $\mathbf{x}$ con respecto a la i-ésima regla disparada.

3. Aprendizaje de reglas difusas lingüísticas ponderadas mediante WCOR

Aunque existen diferentes aproximaciones al problema del aprendizaje de RDLs, este trabajo se centra en los llamados métodos de rejilla o cuadrícula. Éstos asumen que las variables lingüísticas se han definido previamente, y se centran en el proceso de generación de reglas.

En este artículo, se considera el modo más fácil (y más frecuentemente usado) para construir la base de datos lingüística: (1) se usa el mismo número de etiquetas lingüísticas (l) para todas las variables; y (2) se crea una partición simétrica difusa del dominio mediante el uso de l conjuntos difusos triangulares (ver Figura 1).

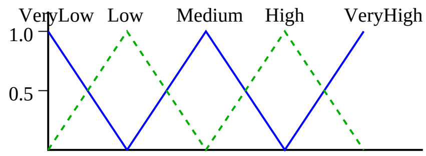


Figura 1. Variable simétrica lingüística con 5 etiquetas

De este modo, tenemos $n + 1$ variables lingüísticas, $\{X_1, \ldots, X_n, Y\}$, en nuestra base de datos lingüística, y cada una tiene asociadas l etiquetas/términos lingüísticos $\{v_i^1, \ldots, v_i^l\}$. El objetivo es aprender la base de reglas.

Los métodos basados en rejilla asumen que todas las variables aparecen en el antecedente de la regla, así que comienzan por definir una cuadrícula n-dimensional $X_1 \times X_2 \times \cdots \times X_n$ donde cada celda o subespacio representa un posible antecedente. Aunque esto da lugar a un número máximo de l^n reglas, estos métodos son guiados por un criterio de cobertura de los ejemplos del *conjunto de entrenamiento*, de modo que los subespacios vacios se descartan. Un ejemplo $e_r = (\mathbf{x}_r, y_r) = (x_{r1}, \ldots, x_{rn}, y_r)$ pertenece a un subespacio S_i si $\forall_{i=1..n} \mu_{S_{ij}}(x_i) > 0$, siendo S_{ij} la variable lingüistica asociada a la variable X_j en el subespacio S_i. Hay que observar que, como los términos lingüísticos se solapan (ver fig. 1) el mismo ejemplo puede pertenecer a distintos subespacios. Sin embargo, la mayoría de los algoritmos basados en cuadrículas para aprender LFRBS asignan un ejemplo e_r a un único subsepacio S_i (el que tiene mayor grado de cobertura).

El siguiente paso consiste en identificar el consecuente para cada subespacio no vacío. Dado un conjunto de ejemplos $e_{S_i} = \{e_1, \ldots, e_m\}$ cubiertos por el subespacio S_i, el algoritmo de Wang and Mendel (WM)[9], que es el más representativo y eficiente de este tipo de métodos, decide, de manera voraz (máxima cobertura e independientemente del resto de reglas), el consecuente para S_i como $Y = v_y^b$, tal que,

$$v_y^b = \arg_k \max_{r=1,\ldots,m} \left[\left(\max_{k=1,\ldots,l} \mu_{v_y^k}(y_r) \right) \prod \mu_{S_{ij}}(x_{rj}) \right]$$

En [3,4] se proponen las metodologías COR y WCOR (*(Weigthed) COoperative Rules*), un método basado en el anterior que trata de superar los inconvenientes del algoritmo WM mediante el estudio de la cooperación de las diferentes reglas del sistema. Así, en la metodología WCOR, la selección voraz de cada consequente para cada subespacio se reemplaza por una búsqueda local en el espacio de todos los conjuntos de reglas candidatos. Para convertir el problema en uno de optimización combinatoria, se necesita definir el espacio de búsqueda de modo que los puntos o individuos del espacio sean evaluados. Una vez que estos componentes han sido especificados se pueden obtener distintas instancias de WCOR mediante el uso de metaheurísticas.

El espacio de búsqueda WCOR: Sea $\{S_1, \ldots, S_m\}$ el conjunto de subespacios no vacios con el criterio de cobertura descrito, el objetivo es buscar los consequentes $\{c_1, \ldots, c_m\}$ que den lugar al mejor sistema posible (ver abajo la definición de la función de fitness).

La definición del espacio de búsqueda es el producto cartesiano: $\{v_y^1, \ldots, v_y^l\}^m$. Sin embargo, WCOR también usa un criterio de cobertura del conjunto de entrenamiento para restringir el número de posibles consecuentes para cada subespacio: Dado un conjunto de ejemplos $\{e_1, \ldots, e_r\}$ cubiertos por un subespacio S_i, entonces el conjunto de posibles consequentes para S_i es:

$$cons(S_i) = \left\{ v_y^k \mid k = \arg \max_{1,\ldots,l} \mu_{v_y^k}(y_r) \right\} \cup \{\aleph\}$$

donde $\aleph$ es un consecuente vacío cuyo significado es que ninguna regla es añadida al sistema para ese subespacio. Además, el proceso de aprendizaje se extiende para inducir no sólo la regla sino su peso. Por ello, es necesario tratar un problema híbrido,

es decir, de optimización numérica más combinatoria. Si hay m posibles consecuentes, el espacio de búsqueda para WCOR es:

$$(cons(S_i), w \in [0,1])^m$$

aunque por conveniencia es mejor situar los enteros (y flotantes) de manera consecutiva:

$$cons(S_1) \times cons(S_2) \times \cdots \times cons(S_m) \times [0,1]^m.$$

Por tanto, un individuo del espacio de búsqueda WCOR es un array $cw[]$ de tamaño $2m$ donde las primeras m posiciones son el número de posibles subespacios (reglas). Para una posición determinada $1 \leq j \leq m$, $cw[j]$ es un número entre 1 y l representando el índice de término linguístico elegido como consequente para el antecedente S_j, o -1 ($\aleph$) que corresponde al hecho de no incluir ninguna regla para el subespacio S_j en el sistema. Mientras que las últimas m posiciones son números reales que representan los pesos. Además, hay una correspondencia entre ambas partes de modo que las posiciones $1 \leq i \leq m$ y $i + m$ están ligadas y representan el consecuente para la regla asociada con el subespacio $S - i$ y su peso(w_i).

Función de evaluación: Por último, para evaluar la bondad de una solución dada $cw[]$, se decodifica en su correspondiente SBRDL y se usa para predecir la variable de salida para los ejemplos del conjunto de entrenamiento. Entonces, el error cuadrático medio (ECM) (o alguna de sus variantes) se usa como medida de bondad.

4. Algoritmos de Estimación de Distribuciones

Los *Algoritmos de Estimación de Distribuciones* (AEDs) [5] son una familia de algoritmos evolutivos que han ganado importancia en los últimos 5 años. Se basan en poblaciones como los algoritmos genéticos (AGs) pero, en lugar de evolucionar mediante operadores genéticos, reflejan las características de los mejores individuos de una población en una distribución de probabilidad y la usan para muestrear nuevas soluciones. El proceso evolutivo de un AED consta de los siguientes pasos:

1. $D_0 \leftarrow$ Generar la población inicial (m individuos)
2. Evaluar la población D_0
3. $k = 1$
4. Repetir
   a) $D_{tra} \leftarrow$ Seleccionar $n \leq m$ individuos de D_{k-1}
   b) Estimar/aprender un modelo nuevo $\mathcal{M}$ a partir de D_{tra}
   c) $D_{aux} \leftarrow$ Muestrear m individuos de $\mathcal{M}$
   d) Evaluar D_{aux}
   e) $D_k \leftarrow$ Seleccionar m individuos de $D_{k-1} \cup D_{aux}$
   f) $k = k + 1$

   Hasta la condición de parada.

Dependiendo de la complejidad del modelo considerado, ya que es impracticable tratar con la distribución de probabilidad conjunta, surgen diferentes modelos de AEDs ya que, cuanto más complejo es éste, más dependencias entre las variables refleja, pero

también es más costosa su estimación. En la literatura pueden encontrarse varias propuestas que se pueden agrupar en: modelos *univariados* (que no permiten dependencias), *bivariados* (permiten dependencias entre pares), y modelos *n-variados*. En este trabajo, nos centramos en los algoritmos univariados y bivariados ya que ofrecen un buen equilibrio entre complejidad y precisión. Concretamente, usamos los algoritmos UMDA, UMDA_g y MIMIC.

Univariate Marginal Distribution Algorithm *UMDA [10]* : Los algoritmos univariados asumen independencia marginal entre las variables y por tanto la distribución de probabilidad conjunta n-dimensional factoriza como:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i) \tag{2}$$

Es decir, sólamente se requiere estimar las probabilidades marginales durante el aprendizaje de los parámetros y cada variable se puede muestrear a partir de éstas de forma independiente. En el caso de que las variables sean discretas, tan sólo bastaran recolectar las frecuencias de cada variable para realizar algún tipo de estimación (máxima verosimilitud, bayesiana, etc.). En el caso contínuo, UMDA *gaussiano* o (UMDA_g) [11], se usa la distribución normal para modelar la densidad de cada variable, y la densidad conjunta se factoriza como el producto de todas las densidades normales unidimensionales e independientes. Así, el aprendizaje del modelo se reduce a la estimación de la media μ y varianza σ^2 para cada variable a partir de la muestra de ejemplos D_{tra}.

Mutual Information Maximising Input Clustering *MIMIC [12]* : En los modelos bivariados las distribución de probabilidad conjunta n-dimensional es factorizada como:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | pa(x_i)),$$

donde $pa(x_i)$ es la variable a la cual se condiciona x_i. $pa(x_i)$ puede ser vacío, por lo que puede haber variables sin padres. En el algoritmo MIMIC, el modelo probabilístico tiene la forma de una cadena ($X_{\pi_1} \rightarrow X_{\pi_2} \rightarrow \cdots \rightarrow X_{\pi_n}$),donde π es una permutación de las n variables y X_{π_i} el elemento de la permutación en la posición i. De este modo, todos los nodos tienen un padre excepto la raiz de la cadena. El algoritmo MIMIC necesita, por tanto, de un aprendizaje estructural que se lleva a cabo mediante el cálculo de la información mutua entre todos los pares de variables (para más detalles consultar [5]). Posteriormente se estiman los parámetros para cada $P(x_i|pa(x_i))$. En la fase de muestreo se recurre la muestreo lógico probabilístico siguiendo el orden establecido por la cadena aprendida [5].

5. Aproximación a WCOR mediante AEDs

En esta sección, se presentan las dos propuestas para tratar el problema del aprendizaje de SBRDLs. En todos los casos, nuestro punto de partida es la representación de los individuos y la función fitness descrita en la sección 3.

La primera propuesta para aplicar los EDAs al problema wCOR surge como una adaptación directa del algoritmo GA-wCOR descrito en [4]. Por tanto, los individuos se componen de una parte entera y una real.

En primer lugar, hemos implementado dos algoritmos UMDA-wCOR y MIMIC-wCOR, cuyos modelos probabilísticos están también compuestos por dos partes independientes: Una de ellas usada para aprender y muestrear los consecuentes de las reglas y la otra usada para modelar los pesos.

En el algoritmo UMDA-wCOR, no sólamente se supone que las dos partes son independientes, sino que todas las variables del modelo probabilístico lo son, esto es, si observamos la figura 2 el primer modelo propuesto sería sin ningún arco entre las variables. Esto es equivalente a usar la combinación de UMDA y UMDA$_g$. Sin embargo, MIMIC-wCOR usa el algoritmo MIMIC para modelar la parte entera del modelo probabilístico, pero también asume la independencia entre todas las variables que representan los pesos, e independecia entre éstos y sus consecuentes asociados. Si observamos la figura 2, este algoritmo se correspondería con el modelo de la parte derecha pero sin ningún arco entre x_{π_i} y los pesos w_{π_i}.

Hay que darse cuenta que, de este modo, la población es el único nexo entre los consecuentes y los pesos. Esto es, como los dos modelos probabilísticos se aprenden del mismo subconjunto de individuos (la mejor mitad de la población) entonces estamos considerando indirectamente relaciones entre las componentes consecuentes y pesos.

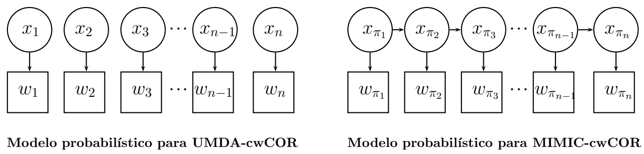


Figura 2. Representción de los modelos probabilísticos usados en UMDA-cwCOR y MIMIC-cwCOR

5.1. Propuesta alternativa: algoritmos de aprendizaje UMDA-cwCOR y MIMIC-cwCOR

Mientras que en los algoritmos previos, los consecuentes y sus pesos son tratados de manera independiente por el modelo probabilístico, en esta sección se proponen dos alternativas en las cuales la relacción entre cada regla y su peso se recoje de manera explícita.

La idea es que el peso asignado a cada regla debería ser claramente dependiente del valor seleccionado como consecuente para esa regla. Esto es, dado un subespacio S_i podemos tener dos posibles reglas *buenas* como “si S_i entonces $Y = c_i$ con $w_i = 0{,}2$” y “si S_i entonces $Y = c_j$ con $w_i = 0{,}9$”, donde claramente el peso es altamente dependiente del consequente. En UMDA-wCOR y MIMIC-wCOR estas dependencias se pierden por que los parámetros (media y varianza) para w_i son estimados sin tener en cuenta el consecuente asociado.

En este trabajo se propone aprender los pesos de manera *condicionada* al valor seleccionado para el consecuente correspondiente. Esto es fácil de hacer en el paradigma AED, por que podemos expresar fácilmente estas dependencias mediante el uso

de un modelo gráfico probabilístico. Así, proponemos usar UMDA-cwCOR y MIMIC-cwCOR cuya estructura gráfica se ve en la Figura 2. Como puede verse, otra vez usamos UMDA y MIMIC para la parte entera pero ahora se usa un modelo probabilístico mixto en lugar de dos modelos separados para reflejar de maner explícita la dependencia entre consecuentes y sus pesos.

Estos nuevos modelos no incrementan la complejidad del aprendizaje estructural (con respecto a UMDA-wCOR y MIMIC-wCOR) por que símplemente se añade un enlace $c_i \rightarrow w_i$ para cada subespacio. El aprendizaje de parámetros es un poco más costoso en espacio (aunque no en tiempo) debido al hecho de aprender una distribución unidimenesional normal condicionada a cada elemento de cons(S_i) en lugar de una simple para cada subespacio S_i.

6. Evaluación experimental

Para llevar a cabo una evaluación experimental de los métodos propuestos, estos han sido testeados con un número de problemas significativos. En esta sección se describen tanto las configuraciones de los algoritmos como los resultados obtenidos.

Conjunto de problemas: Para los experimentos se han usado 3 problemas reales, los dos primeros tomados del repositorio FMLib (`http://decsai.ugr.es/fmlib`).

• Problema $ele1$: Consiste en encontrar un modelo que relacione la *longitud total de una línea de bajo voltaje* instalada en una ciudad rural con el *número de habitantes en la ciudad* y la *media de las distancias desde el centro de la ciudad a los tres clientes más lejanos en ella*. El objetivo es usar el modelo para estimar la longitud total de la línea que ha de ser mantenida.

Por tanto, se tienen dos variables predictivas ($x_1 \in [1, 320]$ y $x_2 \in [60, 1673{,}33]$) y una variable de salida definida en $[80, 7675]$. La cardinalidad de los conjuntos de entrenamiento y test es de 396 y 99 respectivamente.

• Problema $ele2$: En este caso, el modelo trata de predecir los mínimos costes de mantenimiento. Hay 4 variables de entrada: *suma de las longitudes de todas las calles en la ciudad, area total de la ciudad, area total ocupada por los edificios* y *energía proporcionada a la ciudad*.

Los dominios para las 4 variables predictivas son: $[0{,}5, 11]$, $[0{,}15, 8{,}55]$, $[1{,}64, 142{,}5]$ y $[1, 165]$. La variable de salida toma sus valores en $[64{,}47, 8546{,}03]$. Para este problema, el tamaño del conjunto de entrenamiento es de 844, mientras que el número de instancias para el test es de 212.

El problema no perteneciente a FMLib es relativo al campo de la ganadería

• Problema $sheeps$: El marco en que se define el problema es el esquema de selección genética estudiado en Castilla-La Mancha (Spain) con el ánimo de mejorar la producción de leche en la oveja manchega. El principal parámetro de este esquema es el mérito genético de un animal, que es estimado mediante el uso de una metodología estandard (BLUP). Sin embargo, antes de que una oveja llegue a ser madre y la lactación sea controlada, no puede aplicarse (BLUP) y entonces se usa un índice de pedigree (la media aritmética entre el mérito genético de la madre y el padre).

El conjunto de datos usado en esta tarea contiene dos variables predictivas (mérito genético del padre y la madre) y el objetivo es predecir el mérito genético mediante el uso de SBRDL ponderadas en lugar del índice de pedigree. La cardinalidad del conjunto de entrenamiento y test es de 1421 y 711 respectivamente.

Función de evaluación: Como se mencionó en la sección 3, para evaluar la bondad de una solución dada se usa el ECM o alguna de sus variantes. Concretamente, nosotros usamos el *Error Cuadrático Medio Estandarizado* (RECM) para medir el error cometido por el sistema cuando se usa para predecir las instancias del conjunto de entrenamiento. Dado un individuo, c o cw y sus reglas correspondientes F, si $\hat{y}$ es la salida generada por F para una entrada $\mathbf{x}$, mientras que y es la salida verdadera, entonces

$$RMSE(cw) = RMSE(F) = \sqrt{\frac{1}{|\mathbf{D}|}\sum_{i=1}^{|\mathbf{D}|}(\hat{y}_i - y_i)^2}$$

donde $|\mathbf{D}|$ es el número de registros del conjunto de datos.

Está claro que el objetivo es encontrar un sistema con el menor error; sin embargo, como en nuestra implementación se busca maximizar, hemos usado la inversa del RECM como función de evaluación, es decir $fitness(cw) = \frac{1}{RMSE(cw)}$.

Finalmente, y antes de describir la configuración de parámetros usada para los algoritmos, debemos decir que estos han sido escritos en Java y que para la definición y evaluación de los sistemas basados en reglas difusas hemos interactuado con FuzzyJess [13] que etambién está escrito en Java.

Algoritmos y configuraciones: Para los algoritmos basados en WCOR, hemos probado los algoritmos descritos en la sección 5. La mayoría de los parámetros son comunes tanto a AEDs como a AGs: El tamaño de población (*popSize*) se ha fijado a 512 y la población D_k se obtiene de los mejores *popSize*) individuos de $D_{k-1} \cup D_{aux}$, donde D_{aux} es la población generada por muestreo, en el caso de los AEDs, o por la aplicación de los operadores genéticos en el caso de los AGs.

Para los AEDs hemos usados una configuración estandard, es decir, se estima el modelo de la k-ésima generación a partir de los 50 % mejores individuos de la población D_{k-1}.

En el caso de los algoritmos genéticos, los individuos que se cruzan se seleccionan con probabilidad proporcional al rango pero, ya que cada cruce produce ocho inidividuos, sólamente se seleccionan $popSize/8$ parejas.

Con respecto a la condición de parada, cada algoritmo puede evolucionar hasta un máximo de 250 generaciones. Sin embargo, se para antes si no se mejora en el fitness medio de una generación a otra.

En todos los experimentos se han considerado particiones simétricas del dominio real con conjuntos difusos triangulares, con 5 y 7 etiquetas para representar cada variable.

Resultados y análisis: Cada uno de los algoritmos se ha ejecutado 20 veces. El RECM de los modelos obtenidos (entrenamiento $\setminus$ test) y el número medio de reglas y evaluaciones llevadas a cabo se muestran para los casos de 5 y 7 etiquetas en el cuadro 1. Ya

que el criterio optimizado por los algoritmos de búsqueda es el RECM de entrenamiento, se ha marcado esta cifra en negrita. Como puede verse, los resultados (entrenamiento $\setminus$ test) son mejores en casi todos los casos si se usan 7 etiquetas, aunque los modelos obtenidos tienen más reglas y son menos comprensibles. En este punto, es necesario apuntar que el número de reglas no ha sido tomado en cuenta en la función de evaluación y no se ha llevado a cabo nintún tipo de optimización del sistema resultante.

Centrándonos en los algoritmos de búsqueda, que son el objetivo del estudio, está claro que aquellos que usan pesos son más precisos modelando el conjunto de test que aquellos que no. En concreto, puede verse que los algoritmos cwCOR destacan sobre los demás tanto usando 5 como 7 variables. En el primer caso, UMDA-cwCOR parece ser el algoritmo más detacado ya que consigue el menor error en los 3 problemas. Cuando se usan 7 etiquetas por variable, este algoritmo es el más preciso en una ocasion, mientras que en otras dos lo es MIMIC-cwCOR.

Sin embargo, en muchos casos las diferencias entre los resultados son muy pequeñas, por tanto, hemos determinado cuales no tienen diferencia estadística significativa con el mejor mediante tests no pareados de Mann-Whitney. El cuadro 2 muestra los resultados de las comparaciones. Hemos marcado con ○ aquel algoritmo que ofrece el menor error para la predicción de entrenamiento. Después, hemos marcado aquellos que no presentan una diferencia significativa ($p - value > 0{,}05$) con •.

Resultados usando 5 etiquetas para cada variable.

Problem	GA-wCOR	UMDA-wCOR	UMDA-cwCOR	MIMIC-wCOR	MIMIC-cwCOR
ele1	575.0993 \ 610.8644 14.8 – 100844.25	572.2036 \ 608.7981 16.8 – 108828.15	**571.6744** \ 613.3346 17.3 – 84309.7	577.1068 \ 601.3017 15.85 – 38269.5	571.7257 \ 615.8812 17.5 – 71144
ele2	422.2865 \ 428.777 46.1 – 104963.05	371.3274 \ 371.7103 56.1 – 126559.95	**370.2879** \ 376.2354 60.7 – 127806.5	389.618 \ 393.5926 50.15 – 99549.8	370.3089 \ 374.8992 61.9 – 128176.9
ovejas	7.343 \ 6.6741 15.85 – 102100.6	7.2977 \ 6.6475 18 – 54579.95	**7.2878** \ 6.6457 18 – 63533.15	7.3998 \ 6.7304 16.1 – 62385.85	7.2888 \ 6.6424 17.95 – 57466.7

Resultados usando 7 etiquetas por cada variable.

Problem	GA-wCOR	UMDA-wCOR	UMDA-cwCOR	MIMIC-wCOR	MIMIC-cwCOR
ele1	561.3602 \ 670.1182 25.55 – 109533.85	549.1604 \ 672.3695 27.15 – 94108.2	544.4554 \ 694.6266 27.55 – 98417.8	557.6415 \ 677.268 26.05 – 58143.3	**543.4306** \ 695.1677 27.5 – 108671.15
ele2	333.1077 \ 355.9825 84.45 – 109866	258.8464 \ 269.0333 94.05 – 127754.25	254.2318 \ 265.3457 98.35 – 128183.5	279.8552 \ 291.6212 90.85 – 107415.7	**253.723** \ 265.5097 98.2 – 127885.5
ovejas	7.2001 \ 6.9302 31.2 – 114364.65	7.0778 \ 6.8455 33.6 – 115882.4	**7.067** \ 6.8234 34.35 – 113290.45	7.1574 \ 6.9095 31.75 – 72315.6	7.068 \ 6.8306 34.65 – 116548.6

Cuadro 1. Errores promedio en Entrenamiento y Test, número promedio de reglas y número de evaluaciones para 20 ejecuaciones

Mientra que este estudio puede ser visto como una evaluación del análisis de los algoritmos de búsqueda, es necesario estudiar los errores de test para poder evaluar la capacidad obtenida de los SBRDLs. En este caso, el algoritmo de referencia ha sido marcado con □ mientras que los algoritmos que no muestran diferencia significativa han sido marcados con ■ (Cuadro 2). A partir del análisis estadístico, en el caso de entrenamiento, UMDA-cwCOR y MIMIC-cwCOR son los algoritmos que destacan en 2 de los 3 problemas.

Con respecto a los errores de test, los resultados varían dependiendo de que se usen 5 o 7 etiquetas. En el primer caso, ninguno de los algoritmos cwCOR está en el grupo destacado para el problema ele1. Además, UMDA-wCOR y GA-wCOR igualan los resultados para el problema de las ovejas. En el caso de 7 etiquetas por variable, la

Tests para 5 etiquetas para cada variable

Problem	wGA	wUMDA	cwUMDA	wMIMIC	cwMIMIC
ele1		●	○	□	●
ele2		□	○ ■		● ■
ovejas	■	■	○ ■		● □

Tests para 7 etiquetas para cada variable

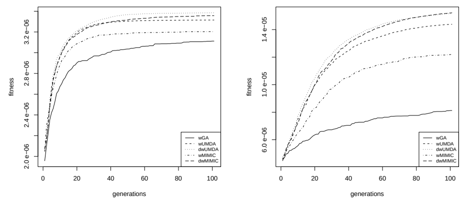
Problem	wGA	wUMDA	cwUMDA	wMIMIC	cwMIMIC
ele1	□	■	●	■	○
ele2			● □		○ ■
ovejas			○ □		● ■

Cuadro 2. Tests estadísticos para entrenamiento y test

tendencia es la misma que en los resultados de entrenamiento (los algoritmos cwCOR destacan sobre los otros) salvo para el problema ele1. Respecto a estos datos, hay que destacar que los modelos obtenidos con 7 etiquetas tienen más precisión al evaluar el conjunto de test que los que usan 5. Además, respecto al problema ele1, y a la vista de los datos obtenidos en el conjunto de entrenamiento, se detecta un problema de sobreajuste de los algoritmos cwCOR. Parece que GA-wCOR y MIMIC-wCOR son claramente peor que los otros algoritmos que usan pesos.

Para ver si estos resultados se deben a una convergencia prematura, hemos representado en la Figura 3 la evolución (media de 10 ejecuciones) del mejor fitness a través de las generaciones para los problemas ele1 (7 etiquetas) y ele2 (5 etiquetas). Como podemos ver, los pobres resultados de estos dos algoritmos no se deben a convergencia prematura. De hecho, UMDA-cwCOR incluso converge antes.

Los resultados obtenidos nos llevan a pensar que usar pesos condicionados con la aproximación wCOR es la mejor opción. El algoritmo UMDA-cwCOR, a pesar de su simplicidad, es casi siempre el mejor o está entre los mejores algoritmos. Además, las gráficas muestran que, dando el mejor resultado, incluso converge con más celeridad.



ele1 (7 labels) ele2(5 labels)

Figura 3. Fitness vs Generación para los algoritmos wCOR

7. Conclusiones y trabajo futuro

En este trabajo se ha hecho una primera aproximación al uso de AEDs como algoritmo de búsqueda en el aprendizaje de SBRDLs poneradas mediante la metodología WCOR. El uso de modelos probabilísticos como elemento principal del proceso evolutivo permite una estimación más precisa de los pesos, lo que da lugar a modelos más precisos. Pensamos que los puntos más destacables de este artículo son: (1) el modo en que se puede sacar ventaja del lenguaje gráfico usado por los AEDs para incorporar el conocimiento del dominio y, (2) el análisis experimental de la metodología WCOR llevado a cabo, que usa 3 problemas reales diferentes mientras que previos análisis están basados en un único problema.

Como trabajo futuro, pensamos extender nuestra investigación en varias direcciones: (1) Una vez que la aplicabiliad de los AEDs a la metodología WCOR ha sido mostrada,

pensamos que debería ser incorporado algún mecanismo de prevención de sobreajuste al proceso de búsqueda. (2) menos reglas conlleva una mayor interpretabilidad, así que planeamos estudiar como reducir el número de reglas en los sistemas inferidos pero sin degradar su rendimiento. Quizá, el problema puede ser enfocado como multiobjetivo considerando el número de reglas y el error como medidas de fitness; (3) También puede ser interesante estudiar el uso de modelos más complejos de EDAs o diferentes métodos para usar la información del dominio.

Agradecimiento

Este trabajo ha sido financiado parcialmente por la Consejería de Educación y Ciencia (JCCM) mediante el proyecto PBI-05-022.

Referencias

1. O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic fuzzy systems: Evolutionary tuning and learning of fuzzy knowledge bases*, World Scientific, 2001.
2. E.H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *Intern. Journal of Man-Machine Studies*, vol. 7, pp. 1–13, 1975.
3. J. Casillas, O. Cordón, and F. Herrera, "COR: A methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules," *IEEE Trans. on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 32, pp. 526–537, 2002.
4. R. Alcalá, J. Casillas, O. Cordón, and F. Herrera, "Improving simple linguistic fuzzy models by means of the weighted COR methodology," in *Advances in Artificial Intelligence -(F.J. Garijo, J.C. Riquelme, M. Toro, Eds.)*. 2002, pp. 294–302, Springer Velarg.
5. P. Larrañaga and J.A. Lozano, *Estimation of distribution algorithms. A new tool for evolutionary computation*, Kluwer, 2001.
6. L.A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 3, no. 1, pp. 28–44, 1973.
7. L.A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.
8. L.A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning," *Information Science*, vol. 8, pp. 199–249, 1975.
9. L.X. Wang and J.M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 22, no. 6, pp. 1414–1427, 1992.
10. H. Mühlenbein, "The equation for response to selection and its use for prediction," *Evolutionary Computation*, vol. 5, pp. 303–346, 1998.
11. P. Larrañaga, R. Etxeberria, J.A. Lozano, and J.M. Peña, "Optimization by learning and simulation of Bayesian and Gaussian networks," Tech. Rep. EHU-KZAA-IK-4-99, UPV, 1999.
12. J.S. de Bonet, C. L. Isbell, and P. Viola, "MIMIC: Finding optima by estimating probability densities," *Advances in Neural Information Processing Systems*, vol. Vol. 9, 1997.
13. R. Orchand, "Fuzzy reasoning in Jess: The FuzzyJ toolkit and FuzzyJess," in *Proc. 3rd International Conference on Enterprise Information Systems*, 2001, pp. 533–542.
14. J. Flores, J.A. Gámez and and J.M. Puerta, "Learning linguistic fuzzy rules by using Estimation of Distribution Algorithm as the search engine in the COR methodology," in *Studies in Fuzziness and Soft Computing, vol. 192*, 2005, pp. 259-280, Springer.

Sociedad híbrida: Una extensión de computación evolutiva interactiva

Juan Romero[1], Penousal Machado[2], Antonino Santos[1] y Marisa Santos[1]

[1] RNASA Lab., Fac. de Informática, Universidade da Coruña, España
{jj, nino, mhyso}@udc.es
[2] CISUC- Centre for Informatics and Systems, Universidade de Coimbra, Portugal
{machado}@dei.uc.pt

Resumen. La falta de un contexto social es un inconveniente en los sistemas actuales de Computación Evolutiva Interactiva. En áreas de aplicación donde las características culturales son particularmente importantes, tales como arte visual y música, este problema es más apremiante. Para solucionar esto, presentamos una extensión del paradigma tradicional de Computación Evolutiva Interactiva que incorpora usuarios y sistemas en un modelo de Sociedad Híbrida, permitiendo la interacción entre múltiples usuarios y sistemas y promoviendo la cooperación. Los resultados de los experimentos, obtenidos con una versión simplificada, validan los mecanismos individuales del modelo propuesto y muestran su capacidad para integrar varios usuarios y sistemas con una relación $n - m$ entre ellos.

1 Introducción

La Computación Evolutiva Interactiva (CEI) es una variación de la Computación Evolutiva en la cual el fitness del individuo se determina mediante una evaluación subjetiva realizada por un usuario humano. En años recientes, este paradigma se ha aplicado a diversos campos, muchos de ellos con un alto componente social, por ejemplo, aquellos relacionados con la estética tales como arte, música, diseño, arquitectura, etc. [1; 2; 3; 4]. Tales domininos se conocen como *dominios sociales*. En ellos podemos distinguir dos roles: el *creador* y el *crítico*. El sistema de CEI canónico integra sólo dos participantes: un sistema de computación evolutiva (CE) y un usuario humano. El rol del creador lo juega el sistema CE, mientras que el rol del crítico lo asume el humano que asigna fitness a los *productos* generados. Los productos pueden ser cualquier tipo de artefactos (ideas, piezas de arte, información, soluciones, etc.). El fitness de los creadores depende del gusto del usuario.

Debido a su naturaleza, estos dominios presentan una serie de características [5] que complican el diseño de los sistemas CEI, tales como:

- La existencia de diferentes funciones de fitness.
- El carácter dinámico del fitness.
- La dificultad de usar formalismos para definir el fitness [6].

El uso de sistemas CEI en dominios sociales plantea las siguientes cuestiones:

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 286-297, 2006.

1. La necesidad de un entorno cultural como el existente en las sociedades humanas; "El valor de una pieza de arte depende de su contexto cultural circundante" [6].
2. La fatiga de usuario causada por la necesidad de evaluar un gran número de individuos.
3. La falta de cooperación entre usuarios. Los resultados obtenidos por ciertos usuarios no son compartidos con otros.
4. La falta de un interfaz común. El usuario debe interaccionar con cada sistema CEI independientemente, usando diferentes interfaces.
5. La capacidad de evaluación del sistema. Los sistemas CEI canónicos carecen de la capacidad de tomar decisiones respecto a la evaluación de sus salidas, lo que fuerza al usuario a evaluar la población entera.

Estos problemas están relacionados con la relación $1-1$ establecida en los sistemas de CEI canónicos entre el usuario y el sistema y pueden ser aliviados estableciendo un modelo con una relación $n-m$. Esta relación $1-1$ también dificulta la integración del sistema CEI en otros sistemas, y la creación de sistemas complejos que incorporen sistemas CEI.

Algunos autores [1; 2; 4; 5] han remarcado la necesidad de entornos comunes que permitan la interacción y validación de diversos sistemas CEI. Un modelo que permita la integración de múltiples sistemas CEI y usuarios, estableciendo una relación $n-m$, puede ser de gran valor para la comunidad investigadora, proporcionando una forma de validar y comparar sistemas CEI, y fomentando la cooperación entre investigadores mediante un marco de trabajo común.

Nosotros proponemos un nuevo paradigma, llamado Sociedad Híbrida (SH), que extiende las habilidades de la actual CEI estableciendo relaciones $n-m$ entre los sistemas de computación evolutiva y los usuarios, así como la incorporación de creadores humanos y críticos artificiales. Comenzamos haciendo un análisis de las variaciones existentes del paradigma CEI. Después, hacemos una descripción conceptual de la arquitectura SH, de sus principales mecanismos, y de algunos detalles de implementación relevantes. En la sección 4, presentamos y analizamos los resultados obtenidos en los experimentos diseñados para evaluar la validez del modelo SH, y la capacidad de permitir una interacción simultánea provechosa entre los usuarios y los sistemas. Finalmente se exponen algunas conclusiones globales.

2 Paradigma CEI

En esta sección analizamos las extensiones actuales del sistema CEI canónico.

2.1 Multiusuario

Una de las variaciones más comunes de la CEI es la aproximación multiusuario, en la que la evaluación de los productos la realizan un conjunto de usuarios en lugar de uno. Típicamente (ver p.ej. [7]) el fitness de cada producto se determina mediante la media de las evaluaciones hechas por los usuarios. Esta aproximación tiene un severo

inconveniente: cuando los usuarios tienen diferentes preferencias, los resultados apenas serán satisfactorios.

En dominios sociales, que implican un alto criterio subjetivo, este problema se acentúa. Además, puede llevar a una "dictadura de la mayoría" donde las preferencias de los grupos minoritarios nunca se vean satisfechas.

2.2 Paralelo

La variante CEI paralela se caracteriza por el uso de un conjunto de sistemas CEI cuyos sistemas CE intercambian individuos de la población. Esta aproximación tiene un problema en campos donde el fitness se asigna de acuerdo con criterios subjetivos tales como el gusto individual. En este tipo de dominios, sería necesario integrar un mecanismo que maximice la migración de productos entre los sistemas CEI de usuarios con preferencias similares, y que minimice las transferencias entre los que tienen gustos diferentes u opuestos. Además, también tiene otro inconveniente, ya que sólo es útil cuando los sistemas CEI usan la misma representación para los individuos. Para usar diferentes sistemas CEI, se necesitaría idear una forma de trasladar los productos de un sistema al otro. Probablemente debido a esto, no hemos sido capaces de encontrar ejemplos de sistemas CEI paralelos aplicados a dominios sociales.

2.3 Parcialmente Interactivo

Otra extensión del paradigma CEI consiste en la integración de mecanismos de evaluación. Estos pueden tener dos objetivos diferentes: llevar a cabo algún tipo de tarea de evaluación que simplifique el trabajo del usuario, por ejemplo, eliminar productos que son inválidos; predecir las evaluaciones del usuario, permitiendo así que el sistema se ejecute en modo "stand alone". En ambos casos la integración de mecanismos de evaluación automática puede contribuir a la disminución de la fatiga de usuario y a incrementar la calidad del resultado global.

El sistema parcialmente interactivo de CE puede tener una capa de filtro para los productos, y las evaluaciones de los mismos pueden ser efecturadas por un Crítico Artificial (CA) o por un usuario humano. En [8] los autores describen un sistema evolutivo parcialmente interactivo, que integra ambos componentes. Una capa de filtro elimina imágenes consideradas claramente insatisfactorias. El CA asigna fitness a las imágenes restantes. El usuario puede interferir en cualquier punto del proceso evolutivo dando su propia evaluación a la población de imágenes, anulando así las valoraciones del CA.

Otra línea activa de investigación se ocupa del desarrollo independiente de CAs. La construcción de una arquitectura genérica que permita una fácil integración de varios CAs y CEs puede ser de gran interés para el desarrollo de sistemas complejos, permitiendo la comparación de diferentes aproximaciones, y fomentando la colaboración entre grupos de investigación.

3 Sociedad híbrida

Teniendo presente los problemas existentes en las diferentes variantes CEI, esta sección describe la Sociedad Híbrida, una extensión del paradigma CEI.

El diseño de esta extensión se basa en los siguientes objetivos:

1. Permitir que cada usuario interaccione simultáneamente con varios sistemas CE.
2. Permitir que cada sistema CE interaccione con diferentes usuarios.
3. Permitir la participación de CAs genéricos que evalúen los productos creados por diferentes sistemas CE.
4. Proporcionar una forma de evaluar el rendimiento de los sistemas CE de acuerdo a su capacidad para satisfacer las preferencias de un conjunto de usuarios y su habilidad para adaptarse a los cambios en estas preferencias.
5. Permitir un mayor grado de interacción entre participantes con preferencias similares, fomentando el desarrollo de grupos con gustos comunes.
6. Simular algunos aspectos del comportamiento de la sociedad humana.

SH fue diseñada específicamente para dominios sociales y está, por lo tanto, basada en una "concepción social". Según esta visión, sólo aquellos productos encontrados interesantes por un *entorno cultural*, son valorados. Un entorno cultural se puede definir como un conjunto de personas con un grado alto de afinidad cultural. Esta concepción no descarta las tendencias de la minoría. Si un trabajo particular de arte provoca el interés de una comunidad pequeña, será valorado. Por ejemplo, el jazz no es un tipo de música de masas, sin embargo, en SH los participantes a los que les gusta el jazz, y los creadores y críticos artificiales que se han adaptado a ese estilo, se agrupan juntos en un subgrupo, posiblemente próspero.

Ahora describiremos la arquitectura de SH, sus principales mecanismos y detalles específicos de su implementación.

3.1 Arquitectura

La arquitectura SH consiste en un elemento central llamado *escenario*, junto con un conjunto de participantes que se comunican con él. Los participantes pueden ser artificiales o humanos, y pueden jugar los roles de creador o crítico.

Conceptualmente, un escenario es el "terreno" común de los seres que participan en una sociedad. Incluye las reglas del juego y define los principios de comunicación entre estos seres. Formalmente, es el conjunto de aplicaciones (bases de datos, protocolos de comunicación e interfaces) con el que interaccionan los creadores, críticos y productos.

En SH los creadores artificiales son instancias de un sistema CE (CES) similar al usado en la CEI estándar. Los críticos artificiales son instancias de un sistema CA (CAS), similar al descrito en la sección 2.3.

Los creadores, humanos o artificiales, envían productos al escenario. Los críticos llevan a cabo evaluaciones de los productos que pertenecen al escenario, comunicando estas evaluaciones por medio de apuestas. La naturaleza humana o artificial de un participante es oculta para el resto. La Figura 1 muestra las relaciones entre los diferentes tipos de participantes y el escenario.

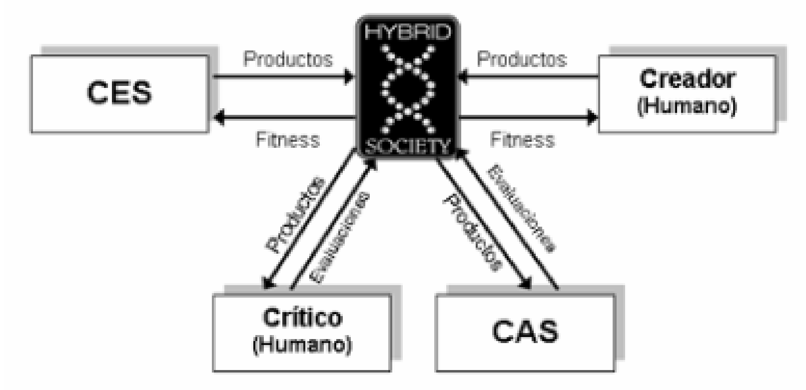


Fig. 1. Modelo de Sociedad Hídrida

Puesto que ya no hay una comunicación directa entre un usuario y un sistema CE, se debe poner en marcha una serie de mecanismos para regular y controlar la integración de todos ellos. Estos mecanismos son: intercambio de energía, afinidad, y generación de descendencia. Las siguientes secciones explican brevemente estos mecanismos, así como las principales variables relacionadas con ellos.

3.2 Intercambio de energía

El mecanismo de intercambio de energía hace posible determinar la adaptación de los participantes al contexto cultural. Cada participante tiene una energía que es una medida de su éxito. Uno de los parámetros de SH es el valor inicial de energía de cada participante. Si la energía de un participante llega a ser menor o igual que 0, entonces está virtualmente muerto y ya no puede participar más en la sociedad.

Los participantes humanos y artificiales ganan y pierden energía según las mismas reglas. Se aplican diferentes reglas dependiendo de si se es creador o crítico.

Cada vez que un creador envía un producto al escenario, una cierta cantidad de energía es sustraida. El creador recibe energía cuando otros participantes hacen apuestas sobre sus productos. Los críticos pueden evaluar los productos y realizar apuestas por los creadores de los que ellos consideran interesantes. Una apuesta es una transferencia de energía de un crítico a un creador. El valor de la apuesta debe ser siempre positiva y menor que la energía actual del crítico que realiza la apuesta. Una vez hechas, todas las apuestas son irrevocables. Cuando un crítico apuesta sobre un creador recibe a cambio un *porcentaje de posesión* de ese creador. El porcentaje de posesión se define como el ratio entre el valor de la apuesta y la energía del creador cuando la apuesta fue realizada.

Cada vez que una apuesta tiene lugar se realizan las siguientes acciones:

1. El valor de la apuesta es sustraido de la energía del crítico que la realiza.
2. Un porcentaje del valor de la apuesta es distribuido entre los apostadores anteriores (si hay alguno) en proporción a su porcentaje de posesión. Cada uno de estos críticos recibe energía según la siguiente fórmula:

$$beneficio_i(t) = \frac{porcentaje_de_posesión_i}{\sum_{j=1..m} porcentaje_de_posesión_j} * E(t) * C, \quad (1)$$

donde *E(t)* es el valor de la puesta *realizada* sobre el creador en el instante *t* y *C* es un parámetro ajustable.

3. El valor restante de la apuesta se suma a la energía del creador.

Una apuesta por un creador con una pequeña cantidad de energía puede llegar a ser más beneficiosa para el crítico que una sobre un creador con una gran cantidad de energía. Los creadores que son evaluados por los participantes de la SH, obtienen grandes cantidades de energía, puesto que reciben muchas apuestas. Los creadores que no satisfacen las demandas de la sociedad no reciben energía. El éxito de los críticos depende de su habilidad para realizar apuestas sobre productos interesantes, que otros miembros de la sociedad encuentran atractivos. Si el crítico apuesta por creadores "no interesantes", nunca recuperará la energía usada para la apuesta. Si el crítico apuesta por creadores "interesantes", estos creadores recibirán más tarde una gran cantidad de apuestas. Puesto que el crítico adquirió un porcentaje de posesión, recibirá parte de estas apuestas y así aumentará su energía. Esta descripción puede llevar a pensar que la mejor estrategia para cualquier crítico es apostar por un creador popular. En la práctica, la mejor estrategia es apostar por creadores emergentes, es decir, aquellos creadores que no son valorados actualmente, pero que lo serán en el futuro.

3.3 Afinidad

Estos mecanismos refuerzan las relaciones entre los participantes con una *afinidad* alta. Dos participantes tienen afinidad alta si: comparten las mismas preferencias (afinidad entre críticos); uno evalúa los productos del otro (afinidad entre crítico y creador); los productos creados por ellos son evaluados por los mismos participantes (afinidad entre creadores). Cuando dos críticos tienen un alto grado de afinidad, tendrán acceso a más productos que hayan sido valorados positivamente por el otro.

Para establecer relaciones de afinidad se usan dos mecanismos basados en una representación espacial. El primero hace que sea más probable que los críticos reciban productos que están espacialmente cercanos. El segundo desplaza productos, críticos y creadores según las evaluaciones de los productos realizadas por los críticos. La representación espacial utilizada puede consistir en dos, tres o más dimensiones. Cada participante toma una posición en esta representación, que inicialmente, es aleatoria. Cuando un creador envía un producto, este producto se sitúa en una posición aleatoria en las inmediaciones de la posición actual del creador. La máxima distancia inicial permitida entre el creador y el producto es un parámetro ajustable.

Para fomentar el establecimiento de relaciones de afinidad, la lista de productos de cada crítico se compone de un mayor porcentaje de productos que están dentro de su vecindad que fuera. Estos porcentajes y la distancia máxima que define la zona de vecindad son también parámetros ajustables. Los participantes y productos afines se acercan de acuerdo a las evaluaciones de los críticos. Cuando un crítico emite una

evaluación positiva de un creador, tienen lugar tres movimientos: (i) el crítico se mueve hacia el producto, (ii) el producto se acerca al crítico, (iii) el creador del producto se mueve hacia el crítico. La distancia máxima por movimiento para creadores, críticos y productos se establece mediante parámetros ajustables independientes. En cambio, también hay fuerzas de rechazo – con una dirección opuesta y de menor magnitud que las anteriores- entre los participantes y productos no afines.

El uso del espacio fomenta la definición de subgrupos dentro de un escenario, dado que la dinámica de SH favorece la proximidad de los participantes que son afines.

3.4 Generación de descendencia

SH ha sido diseñada para que los participantes sean capaces de satisfacer las demandas de una sociedad dinámica. Esto se hace siguiendo una aproximación evolutiva. El mecanismo de generación de descendencia es manejado por SH. Cuando un participante excede un cierto umbral de energía, SH puede crear nuevos participantes artificiales que son sus descendientes. Una vez creados, los descendientes tendrán la mitad de la energía de sus padres. Por defecto, la creación de un nuevo descendiente se hace mediante la mutación del código genético del progenitor. Esta información genética codifica un conjunto de parámetros ajustables que permiten cambiar el comportamiento del sistema.

3.5 Implementación

SH fue diseñada con el objetivo de usar este paradigma con un gran número de participantes, humanos y artificiales.

Comenzamos describiendo el funcionamiento del escenario, y después describimos el interfaz con participantes humanos y artificiales.

Escenario

Se implementa como una pieza de software que comunica a los diversos participantes vía servicios web, pero no analiza o procesa los productos de ninguna forma. La implementación es genérica y adaptable a diferentes dominios sociales.

Antes de comenzar una ejecución deben ser establecidos una serie de parámetros que establecen el comportamiento del escenario. Los sistemas artificiales integrados en la sociedad (sistemas CE y CA), proporcionan soporte a los participantes artificiales, y deben ser ejecutados y conectados al servidor. Similarmente, los participantes humanos deben ser incluidos en la ejecución a través de un interfaz. Después de esto, la ejecución se inicia en el servidor, que conecta a cada uno de los participantes, indicando su comienzo.

Las ejecuciones de SH son organizadas en *pulsos*, que son intervalos de tiempo discretos. El intervalo producido entre dos pulsos consecutivos es el parámetro *duración-pulso*, que permite la sincronización de los diversos participantes. Cada vez que se produce un pulso, el escenario:

1. Modifica los valores de energía de cada participante.
2. Realiza las acciones solicitadas por los participantes.

3. Genera una nueva lista de productos para cada crítico.

Interfaz del sistema CE

Proporcionamos un entorno que permite una fácil adaptación del sistema CE (actualmente usado en el paradigma CEI) al paradigma SH. Con el propósito de facilitar esta adaptación, hay un módulo de control genérico que proporciona varios servicios: mecanismos evolutivos, operadores genéticos, combinaciones de selección, manejo de la generación de descendencia, comunicación con el escenario, etc.

Interfaz de usuario

Los participantes humanos interaccionan con SH vía interfaces basados en web. Han sido usados dos tipos de interfaces. El primero es similar al utilizado por la mayor parte de los sistemas CEI, presentando al usuario una lista de productos para su evaluación. El segundo fue diseñado para "Golem Project"[1]. Este proyecto tiene como propósito fomentar la creación y evaluación de servicios web en un entorno que combina el aprendizaje, la competición y el ocio. La evaluación de este software se hace colectivamente según el interés alcanzado.

4 Experimentos: Resultados

Con el objetivo de probar los mecanismos de SH, y de establecer las reglas y parámetros del escenario, se realizaron una serie de experimentos. En particular, presentamos un experimento que prueba el mecanismo de afinidad y dos más que prueban el mecanismo de intercambio de energía.

4.1 Test dimensional

Para ilustrar y configurar el mecanismo de afinidad independientemente, llevamos a cabo un conjunto de experimentos en los cuales participantes artificiales simples, simulando el comportamiento de críticos y productos, son situados en una representación espacial limitada de dos dimensiones. La localización inicial de cada participante es aleatoria. Para simular los diferentes gustos estéticos, críticos y creadores tienen un *número característico* del 1 al 100, correspondiente al "gusto estético" (en críticos) o a la "característica estética" de productos. Este número es aleatorio para los productos, y viene dado en cada experimento para los críticos.

Debido a los mecanismos de afinidad, los productos y críticos deberían estar organizados en clusters que incluyen aquellos elementos similares (los que tienen números característicos cercanos). En cada iteración:

1. Cada crítico tiene una lista de productos construida siguiendo el principio del mecanismo de afinidad.
2. Cada crítico selecciona uno de los productos de su lista, por medio de un mecanismo de ruleta. Los productos que tienen un número característico cercano al del crítico tienen mayor probabilidad de ser seleccionados.

[1] http://www.golemproject.com

3. El producto seleccionado se acerca al crítico, y al revés.
4. El crítico también selecciona otro crítico mediante un mecanismo de ruleta. Cuanto más diferentes son los críticos, mayores posibilidades tienen de ser seleccionados.
5. Se aplica una fuerza de repulsión entre los críticos.

Ahora presentamos los resultados de tres experimentos diferentes. Realizamos 30 ejecuciones independientes de cada uno de los experimentos, usando diferentes posiciones generadas para productos y críticos. El espacio, de dos dimensiones, consta de 54351 puntos de ancho por 35351 de alto. Los parámetros más relevantes de estos tres experimentos se presentan en la Tabla 1.

Variable	Exp. 1	Exp. 2	Exp. 3
Nº críticos	4	3	3
Nº productos	100	15	10
Forma espacio	Rect.	Rect.	Toro
Nº iteraciones	6000	4000	4000
% de vecinos	80	80	100
Radio de vecindad	1000	5000	6000
Aproximación máxima	250	1000	1000

Tabla 1. Parámetros de afinidad

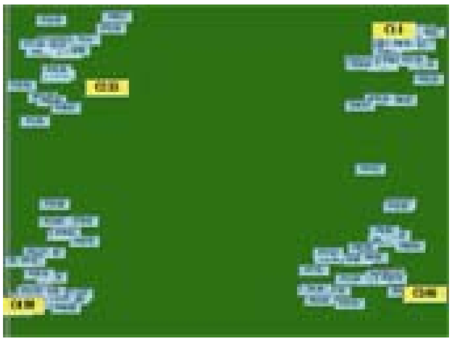


Fig. 2. Posición final de productos y críticos

En el Experimento 1 (Figura 2) cada crítico se sitúa en una esquina y hay una isla de productos similares en la vecindad de cada crítico. En el Experimento 2 se obtienen similares estados. En el Experimento 3 la forma del espacio es un toro, por lo que las islas no están en las esquinas. La Tabla 2 resume los resultados de los experimentos, presentando la distancia media entre cada producto y el crítico más cercano, y la distancia media entre cada crítico y el crítico más cercano a él. Como muestran los resultados, los productos se organizan alrededor del crítico más afín formando islas. Al mismo tiempo, la fuerza de repulsión asegura que los críticos no afines residan en diferentes zonas del espacio.

Experimento	Distancia Media Prod - Crít	Distancia Media Crít - Crít
Experimento 1	4476.2	8536.2
Experimento 2	12722.4	25557.3
Experimento 3	1751.5	13749.5

Tabla 2. Distancia media entre producto y crítico más cercano y distancia media entre crítico y crítico más cercano. Los resultados son medias de una serie de 30 ejecuciones

Estos experimentos indican que el mecanismo de afinidad permite la creación de islas de críticos con preferencias similares y, consecuentemente, de productos interesantes para estos críticos.

4.2 Mecanismo de intercambio de energía

El objetivo es comprobar si el mecanismo de intercambio de energía es suficiente para permitir que las preferencias de los críticos sean transmitidas a los diferentes creadores, permitiendo así su adaptación a estas preferencias. Para este propósito, usamos el Sistema Tribu CE, junto con críticos artificiales que poseen una preferencia estática y concreta. Tribu es un sistema CEI que compone música interactivamente [9]. Las salidas de Tribu son patrones de un compás de diez instrumentos percusivos. Los productos son arrays bidimensionales (16 partes rítmicas de 10 instrumentos) de valores binarios. En el array, “1” significa que el instrumento correspondiente está sonando; mientras que “0” significa un silencio en ese instante de tiempo.

Los críticos no adaptativos usados en estos experimentos evalúan cada producto dependiendo de un algoritmo interno. Cuando al menos un producto tiene una evaluación mejor que un determinado umbral de evaluación (es decir, un parámetro de cada crítico), entonces se realiza una apuesta. Si más de un producto supera dicho umbral, se usa un mecanismo de ruleta teniendo en cuenta la evaluación de cada uno de estos productos. Para el experimento de esta sección se ha usado un CA llamado “OrejaP”, cuyo algorimo interno de evaluación usa una valoración que es proporcional a los silencios de la música recibida. OrejaP realiza apuestas sobre los temas pusicales que tienen un gran número de silencios. Los mecanismos de afinidad no se usan en estos experimentos, es decir, los críticos perciben todos los productos, y no hay reglas de movimiento.

Al comienzo del experimento, había 10 participantes Tribu y 10 OrejaP. Durante las 6500 iteraciones, el porcentaje de silencios de las tribus existentes aumenta continuamente, mostrando que el mecanismo de intercambio de energía es suficiente para permitir la adaptación del sistema CEI a las preferencias de los críticos.

4.3 Mecanismo de intercambio de energía con participantes humanos y artificiales

Una vez que el mecanismo de intercambio de energía fue probado en un entorno simplificado, se llevaron a cabo una serie de pequeños experimentos con participantes humanos y artificiales. El objetivo era comprobar el funcionamiento de este mecanismo con varios creadores y críticos humanos participando simultáneamente.

Los participantes humanos juegan los roles de compositores y críticos. Para esto, usan un simple interfaz web que permite escuchar y crear productos percusivos. Los participantes humanos se sitúan en diferentes habitaciones, sin saber qué participantes son humanos. Varios participantes tribu son usados como creadores artificiales, componiendo temas percusivos mediante técnicas evolutivas. En estos experimentos, los participantes artificiales deben adaptarse a las preferencias de los humanos, en vez de adaptarse simplemente a un criterio simplificado, como en el experimento previo.

Uno de estos experimentos implica a tres humanos (músicos profesionales). Cada participante tribu, aunque realizado por el mismo programa, tiene un conjunto de parámetros diferente, lo que da lugar a diferencias de comportamiento remarcables, provocando diferencias en su adaptación. De 10 participantes tribu, 8 se adaptan a la

sociedad y tienen descendientes (Figura 3). Su energía no cambia significativamente durante los primeros 30 ciclos, y comienza a incrementarse a partir del punto en el que la evolución de los participantes tribu permite la adaptación a las preferencias de los humanos. La mayoría de las apuestas de los humanos van dirigidas a temas producidos por participantes tribu. Esto se puede explicar por la existencia de un mayor número de piezas producidas por ellos. Sin embargo, en términos de número de apuestas medio por pieza, los participantes tribu también obtienen valores medios más altos que los humanos. Aunque la diferencia no es estadísticamente significativa, los resultados exceden nuestras expectativas iniciales.

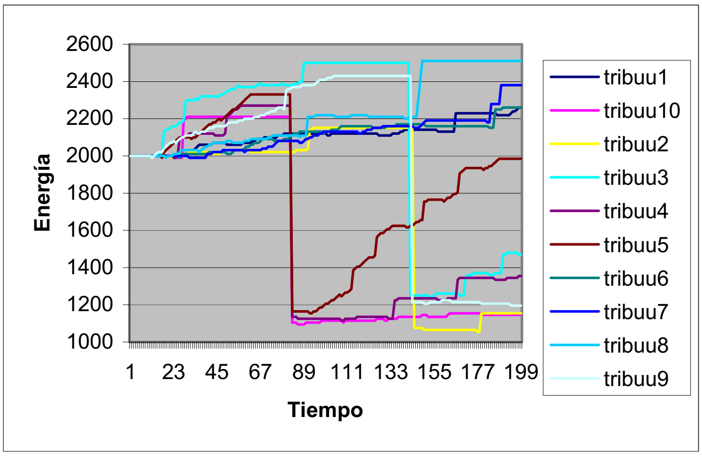


Fig. 3. Energía de los individuos tribu originales

4.4 Análisis de los resultados de los experimentos

Los experimentos presentados validan dos conceptos fundamentales de SH: el establecimiento de relaciones de afinidad y la transferencia de preferencias de los críticos a los creadores, por medio del mecanismo de intercambio de energía.

Los resultados obtenidos indican la adecuación del mecanismo de afinidad, que permite la creación de islas de participantes con intereses similares. Los experimentos referentes al intercambio de energía, en particular el que incorpora críticos humanos y creadores, muestran la capacidad del mecanismo para transferir las preferencias de los críticos a los creadores, permitiendo la adaptación de los creadores a las demandas específicas de la sociedad. Esto se debe a la capacidad de los sistemas CE usados para adaptarse a las preferencias de los usuarios humanos, usando el mecanismo de intercambio de energía como la única herramienta de comunicación entre los creadores y los usuarios. Lo que muestra el experimento es la habilidad de tratar simultáneamente con críticos y creadores con diferentes preferencias estéticas, por medio de este mecanismo.

El conjunto de experimentos llevados a cabo indican la adecuación de los mecanismos del paradigma, quedando patente la viabilidad del modelo de relación n – m.

Resulta, sin embargo, necesario realizar experimentos a una escala más grande (con componente dimensional, participantes humanos y artificiales) para continuar validando el paradigma propuesto.

5 Conclusiones

En este artículo se presenta una extensión del paradigma tradicional CEI llamada Sociedad Híbrida. SH fue diseñada específicamente para dominios sociales, donde la evaluación de los productos depende de criterios subjetivos y culturales. Para superar algunos de los inconvenientes de los sistemas CEI, SH permite la interacción entre una multitud de creadores y críticos, dando lugar a clusters de participantes con relaciones de afinidad altas.

Dada la madurez de la investigación en el área de CEI y sus sólidos logros, consideramos que éste es el momento para un esfuerzo de investigación común, junto con un cambio en las relaciones del paradigma; de una relación 1 – 1 entre usuario y sistema CEI a una relación n – m. Creemos que el uso de SH puede ser un paso importante en el desarrollo de sistemas CEI y CA en dominios sociales y creativos, fomentando la colaboración para su creación y su validación en un entorno dinámico y complejo.

Referencias

1. P. J. Bentley and D. W. Corne, *Creative Evolutionary Systems*, Morgan Kauffmann Publishers Inc, 2001.
2. C. Johnson and J. Romero, "Genetic Algorithms in Visual Art and Music", *Leonardo, 35, 2,* pp. 175-184, 2002.
3. J. Romero, A. Santos, J. Dorado, B. Arcay, and J. Rodriguez, "Evolutionary Computation System for Musical Composition", in *Mathematics and Computers in Modern Science*, World Scientific and Engineering Society Press, pp. 97-102, 2000.
4. P. M. Todd and G. M. Werner, "Frankenstenian Methods for Evolutionary Music Composition", in *Musical Networks: Parallel distributed perception and performance*, MIT Press, Cambridge MA, 1998.
5. A. Pazos, A. Santos, B. Arcay, J. Dorado, J. Romero and J. Rodríguez., "An Application Framework for Building Evolutionary Computer Systems in Music", *Leonardo, 36, 1,* MIT Press, Cambridge MA, pp. 61-64, 2003.
6. P. Machado, J. Romero, B. Manaris, A. Cardoso, and A. Santos, "Power to the Critics – A Framework for the Development of Artificial Art Critics", in *Proceedings of the IJCAI'2003 Workshop on Creative Systems*, pp. 55-64, Acapulco, Mexico, 2003.
7. J. Biles, "GenJam Populi: Training an IGA via audience-mediated performance", in *Proceedings of International Computer Music Conference*, International Computer Music Association, pp. 347-348, 1995.
8. P. Machado, J. Romero, A. Cardoso, and A. Santos, "Partially Interactive evolutionary Artists",
9. A. Pazos, A. Santos, J. Dorado, and J. Romero, "Genetic Music Compositor" in *Proceedings of the 1999 Congress of Evolutionary Computation,* IEEE, 1999.

Vehículos Inteligentes: Aplicación de la visión por computador

Cristina Hilario[1], Juan M. Collado[1], Juan Pablo Carrasco[1], Marco Javier Flores[1], José Manuel Pastor[2], Fº José Rodríguez[1], José Mª Armingol[1], Arturo de la Escalera[1]

[1] Grupo de Sistemas Inteligentes. Dpto. de Ingeniería de Sistemas y Automática. Escuela Politécnica Superior, Univ. Carlos III de Madrid, 28911, Leganés, Madrid.
chilario@ing.uc3m.es
http://www.uc3m.es/islab
[2] Dpto. de Sistemas Informáticos. Escuela Universitaria Politécnica de Cuenca. Univ. de Castilla la Mancha.
josemanuel.pastor@uclm.es

Resumen. En los últimos 20 años, el elevado índice de accidentes en todo el mundo, ha motivado el desarrollo de los llamados vehículos inteligentes. Ámbitos tan diversos como la investigación, la industria automovilística y organizaciones relacionadas con los sistemas de transporte, han aunado fuerzas para incrementar la seguridad vial a través del desarrollo de Sistemas de Ayuda a la Conducción. El objetivo de los Sistemas Inteligentes de Transporte es incrementar la seguridad, eficiencia y confort del transporte mejorando la funcionalidad de los coches y las carreteras, usando las tecnologías de la información. En el presente artículo se presenta al vehículo IVVI (Intelligent Vehicle based on Visual Information). Se trata de una plataforma de investigación para la implementación de sistemas basados en visión por computador, que sirvan de ayuda a la conducción. Se analizan los desarrollos realizados en la detección de señales de tráfico, otros vehículos, peatones y los límites de la carretera.

Palabras Clave: Percepción Artificial, Visión Artificial, Reconocimiento de patrones. Sistemas Inteligentes de Transporte, Visión por Computador, Sistemas de Ayuda a la Conducción

1 Introducción

Los automóviles constituyen el medio de locomoción más utilizado en la actualidad, la congestión del tráfico, el número elevado de accidentes y la contaminación son algunos de los problemas ocasionados. Por otro lado el número de vehículos no para de crecer. En España el parque se ha duplicado en el periodo 1985-2000. Este aumento hace que las infraestructuras se queden pequeñas, provocándose congestiones y falta de fluidez en ellas. El objetivo de los Sistemas Inteligentes de Transporte (SIT) es incrementar la seguridad, eficiencia y confort del transporte mejorando la funcionalidad de los coches y las carreteras, usando las tecnologías de la información.

Existen dos campos en los que se puede trabajar en los SIT: mejoras introducidas en las infraestructuras y dotar a los vehículos de nuevas capacidades. Ambas solucio-

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 298-309, 2006.

nes tienen sus ventajas e inconvenientes. La primera supone una buena opción si se trata de rutas pequeñas y para desplazamientos fijos de vehículos públicos, pero introducirlas en la red viaria total de un país presentaría un coste prohibitivo. Por ello parece más razonable poner el énfasis en los vehículos. Una razón adicional es que el 90% de los accidentes se producen por fallo humano, ya que la mayoría son de día (60%), con buen tiempo (94%), con vehículos en buen estado (98%) y casi la mitad en un trayecto recto (42,8%).

Los campos en los que se está trabajando para lograr la conducción automática son:

- Seguimiento del borde de la carretera.
- Mantenimiento de la distancia de seguridad.
- Regulación de la velocidad dependiendo del estado del tráfico y del tipo de la carretera.
- Adelantamientos de otros vehículos.
- Trazado automático de la ruta más corta.
- Movimiento y aparcamiento dentro de las ciudades.

Los beneficios de una conducción completamente automática son numerosos, pero existen varias dificultades:

- Técnicas. Los algoritmos que realicen estas tareas deben trabajar en tiempo real y tener un grado de fiabilidad del 100%.
- Económicas. Su incorporación no debe suponer un gran incremento del coste actual de los vehículos.
- Psicológicas. Los ocupantes de los vehículos deben acostumbrarse a no controlar la marcha de éstos y ser conducidos por un ordenador.
- Legales. Ante un accidente no estaría claro quién sería el responsable: si el conductor o el fabricante del vehículo.

Por ello, más que en lograr una conducción automática, parece más sensato poner el énfasis en el desarrollo de Sistemas de Ayuda a la Conducción (Driver Assistance Systems). El que al final sean un paso intermedio a ella o no, dependerá sobretodo de cómo se resuelvan los problemas legales y psicológicos. Los equipos deben ser los mismos pero se relaja la exigencia de robustez del sistema y se logra que los conductores vayan confiando en el ordenador mientras se reduce el número de accidentes.

Dentro de estas ayudas a la conducción, los sistemas más importantes para un vehículo son:

- Sistema de aviso en caso de adormecimiento (Drowsy Driving Warning System). Determinan el grado de atención del conductor y le avisan en caso de que esté durmiendo.
- Control de velocidad variable (Adaptive Cruise Control) Se adapta la velocidad a la del vehículo que hay enfrente manteniendo la distancia de seguridad accionando el acelerador y el freno.
- Sistema anti colisión. (Anti Collision Assist) parecido al anterior pero ahora los obstáculos son coches parados, objetos en la vía, etc. Se avisa al usuario de la presencia de obstáculos o coches detenidos o a velocidades muy bajas.
- Parar y marchar (Stop & Go). Mantendrían el control del vehículo a bajas veloci-

dades, por ejemplo en colas para entrar en las autopistas, en semáforos o en peajes.

- Sistema de ángulos muertos (Overtaking Warning) Los sensores cubren el ángulo muerto del vehículo avisando de la presencia de otros coches que estén realizando un adelantamiento.
- Alejamiento del lateral (Lane Departure Warning). El sistema detecta de forma automática la posición respecto a la línea lateral avisando al conductor si la va a sobrepasar de forma inadvertida.

La visión por computador presenta una serie de ventajas frente a otros sensores como radares y láseres:

- La mayoría de los accidentes se producen de día y con buen tiempo (buena visibilidad). No sería necesario por tanto disponer de otros sensores como radares o láseres que funcionan mejor que las cámaras en condiciones ambientales adversas.
- Los radares y láseres detectan solamente los obstáculos que están justo enfrente del vehículo. No pueden percibir por tanto los vehículos que circulan en otros carriles y además pierden al vehículo delantero en las curvas.

2 Vehículo IvvI

La plataforma de investigación IvvI se muestra en la figura 1. Los cuatro sistemas que la conforman son:

- Sistema de posicionamiento. Permite la integración temporal de las observaciones de las diversas cámaras colocadas en el coche.
- Sistema de percepción. Está constituido por cinco cámaras CCDs. En la parte frontal del vehículo está instalada una cámara color para la detección y análisis de la señalización vertical de la carretera, así como dos cámaras B&N, tanto para la detección de obstáculos como la localización de los bordes de la calzada. En la parte posterior se situarán dos cámaras para la percepción de los ángulos muertos.
- Sistema de procesamiento. Está formado por las tarjetas de adquisición de imágenes y una red de ordenadores.
- Sistema interfaz con el conductor. Interacciona con el conductor avisándole sobre la información recogida de la carretera y la conveniencia o no de las maniobras que realiza.

Las capacidades (figura 2) sensoriales que presenta el vehículo son:

- Señalización vertical. Detección de las señales de tráfico y los paneles informativos.
- Vehículos y peatones. Se detectan los diversos objetos que rodean al sistema estimando su velocidad y trayectoria.
- Detección de la carretera. Se detectan los diversos carriles que tiene la carretera.
- Ángulos muertos. Comprobación de la existencia y velocidad de otros vehículos.

- La combinación de estas habilidades da lugar a un análisis más complejo del entorno. Los módulos contemplados son (figura 2):
- Módulo anti-colisión. Tiene en cuenta la posición del vehículo respecto a las líneas laterales de la carretera y los vehículos que le rodean.
- Sistema de supervisión de la velocidad. Indica al conductor la velocidad correcta en función de la propia velocidad del vehículo, la de los que lo rodean y de las señales viarias.
- Módulo de adelantamientos. Evaluará la maniobra en función de las señales de tráfico y líneas de la carretera. Además comprobará la presencia de coches en sentido contrario.



Fig. 1. Vehículo IVVI con sus sistemas.

3 Detección de señales de tráfico

La detección automática de señales de tráfico ha recibido un interés creciente por parte de los laboratorios de investigación. Ello es debido a las aplicaciones que se podrían desarrollar como:

- Mantenimiento de autopistas. Actualmente es un operador el que tiene que observar una cinta de video para determinar si la señal de tráfico está en buen estado y goza de buena visibilidad.
- Inventario de señales en ciudades. En este entorno las señales no están siempre perpendiculares al movimiento del vehículo, hay objetos con el mismo color y las oclusiones son más frecuentes.
- Sistemas de ayuda a la conducción. Su interpretación facilita supervisar la velocidad, y la trayectoria del vehículo.

Las principales dificultades (figura 3) son:

- Las condiciones de iluminación son cambiantes y no controlables.

- La presencia de otros objetos da lugar a oclusiones y sombras.
- El rango de posibles variaciones de la apariencia del objeto en la imagen es muy grande.

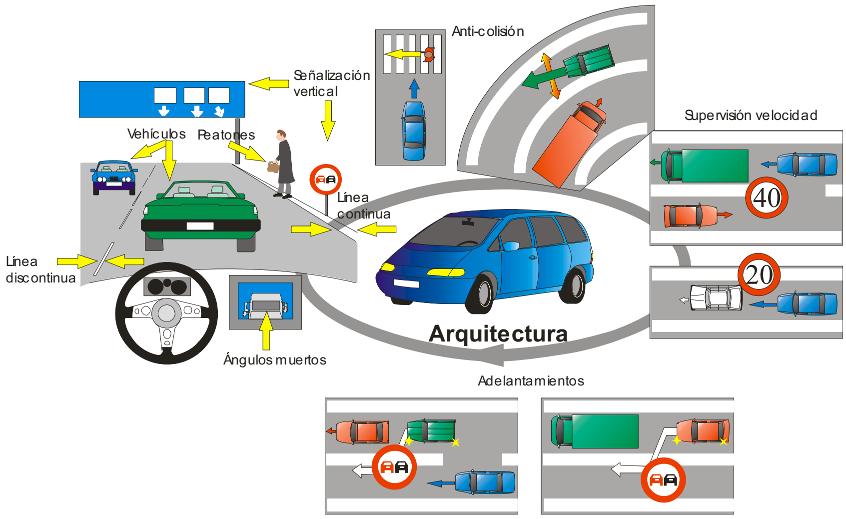


Fig. 2. Capacidades sensoriales y módulos del vehículo IVVI.

La detección de señales de tráfico puede realizarse analizando imágenes en color o en niveles de gris. Dentro del primer grupo se han realizado trabajos con los espacios de color estándar como RGB [7]. Debido a los conocidos problemas de iluminación también se ha utilizado el espacio HSI y el Luv . Estudios más exhaustivos se han realizado construyendo una base de datos o con el uso alternativo de texturas, clasificadores borrosos o redes neuronales. Ninguno considera que la clasificación de los píxeles pueda ser errónea lo que, como se verá más tarde, es una limitación. Los trabajos que parten directamente de una imagen en niveles de gris realizan una detección de bordes que analizan más tarde buscando la forma de la señales. Así, en algunos casos se utiliza una estructura piramidal, en otros algoritmos genéticos y en [11] templado simulado.

El algoritmo propuesto en este artículo parte de un análisis del color para determinar si existen zonas en la imagen donde pueda existir una señal. Si así ocurriese se buscan siguiendo un modelo deformable que tiene en cuenta el color, y los bordes. Para encontrar la instancia concreta en la imagen se utilizan Algoritmos Genéticos (AG) [8].

El análisis del color es fundamental ya que el diseño de las señales de tráfico ha sido realizado teniendo en cuenta esta característica. Debido a los problemas de iluminación el espacio de colores utilizado es el HSI. Solo las dos primeras componentes van a ser empleadas. Para ello se construyen dos Tablas de Consulta por cada color buscado, con la finalidad de resaltarlo a pesar de los cambios en la iluminación (figura 4). Como puede observarse pueden existir otros objetos con el mismo color y no siempre se van a poder clasificar todos los puntos correctamente.



Fig. 3. Principales dificultades para detectar las señales de tráfico

Una imagen concreta de una señal de tráfico presenta los siguientes grados de libertad: posición en la imagen y escalas diferentes en cada eje de la imagen ya que la señal no va a estar siempre perpendicular al eje óptico de la cámara, ni a la misma distancia. Estos grados pueden expresarse matemáticamente mediante una transformación afín de la imagen de una señal situada a una distancia determinada y por cuyo centro pasa el eje óptico de la cámara.



Fig. 4. Realce del color

El algoritmo de búsqueda de los valores de esa transformación que mejor se ajusten a la señal presente en la imagen se realiza mediante AGs, considerando el equilibrio que presentan entre las tareas de exploración y explotación. Un ejemplo de búsqueda se observa en la figura 5.



Fig. 5. Ejemplo de búsqueda de una señal en una imagen

La función que mide lo bien que un modelo concreto se ajusta a la imagen se basa en la distancia de Hausdorff, que mide la separación entre dos conjuntos de puntos. En el presente caso los bordes del modelo deformado y los de la imagen de color. Los ejemplos de las señales detectadas se muestran en la figura 6. Para el reconocimiento se ha utilizado una red neuronal de tipo ART1 ya que es capaz de almacenar el conocimiento y no necesita ser reentrenada si se presentan nuevos tipos, por lo que el funcionamiento de la red le sirve a su vez de entrenamiento.



Fig. 6. Ejemplos de detección de señales

4 Detección de vehículos

La detección de vehículos es fundamental para las siguientes tareas:

- Seguimiento en pelotón. Los vehículos circulan en grupo a altas velocidades y con distancias de separación pequeñas.
- Stop&go. Básicamente es lo mismo que el problema anterior pero para el caso de conducción dentro de una ciudad.
- Ángulo muerto. El sistema tiene que detectar que hay otro vehículo aproximándose lo que impediría que el nuestro iniciase un adelantamiento.
- Supervisor de maniobras propias y las de los demás vehículos.

Aunque ahora mismo se utilizan sensores distintos a las cámaras [23] [28], ya se han comentado con anterioridad los inconvenientes de los láseres y radares. Los enfoques basados en visión se pueden clasificar en tres grupos:

- Por características. Así la sombra inferior es utilizada en algunos trabajos, movimiento en [1]. La simetría de los niveles de gris es también empleada, mientras que otra opción es la simetría entre bordes verticales [10]. También se puede aplicar los bordes horizontales [20] o una suma ponderada de todos [3]. Estas características se van buscando sucesivamente hasta llegar al vehículo. Tienen el inconveniente de ser decisiones todo o nada en la que la ausencia de una de ellas imposibilitaría detectar a los vehículos.
- Modelos. Se parte del modelo (o modelos) del vehículo que se buscan en la imagen. Es más robusto que el anterior pero suele llevar más tiempo. Pueden ser 2D [13] o 3D [9].
- Aprendizaje. Mediante el cálculo de las distribuciones de probabilidades de los niveles de gris [18] o por redes neuronales [26].

El algoritmo desarrollado se engloba dentro del grupo que define un modelo del vehículo y busca en la imagen aquella zona que mejor se ajuste.

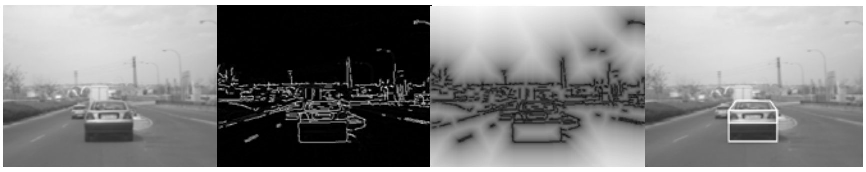


Fig. 7. Algoritmo de detección de vehículos

A diferencia del caso de las señales de tráfico, ni el color ni el nivel de gris definen todas las disposiciones que puede tener un coche. Es por ello que se ha tomado solo la forma. Los parámetros que la definen son: altura y anchura del vehículo, altura del parabrisas, altura del maletero e inclinación del techo [15] [17]. Junto a la posición, forman los siete parámetros que definen el modelo. De nuevo se han utilizado algoritmos genéticos para la búsqueda, siendo la distancia de Haussdorf la que indica lo bien que se ajusta un modelo concreto a la imagen. Los resultados pueden verse en la figura 7.

5 Detección de peatones

La protección de los elementos más vulnerables de la circulación, ha recibido muy poca atención a la hora de desarrollar vehículos inteligentes. El hecho de que el entorno de trabajo sea exterior, la gran variación de apariencia y movimientos que pueden presentar las personas, así como el movimiento de la cámara instalada en el vehículo, hacen que el desarrollo de un sistema de detección de peatones sea complicado. Por este motivo, son pocos los vehículos que en la actualidad integran un sistema de este tipo.

Los métodos de sustracción de imágenes, tradicionalmente empleados en aplicaciones de vigilancia, no pueden usarse por el movimiento del vehículo. Por otro lado, el flujo óptico es difícil de aplicar debido al movimiento no-rígido de las personas. Tampoco son adecuadas otras técnicas de segmentación, como las basadas en la intensidad, ya que las condiciones de iluminación son cambiantes.

Para el caso de aplicaciones sin restricciones, los métodos estadísticos son más eficaces dada su adaptabilidad. Sin embargo, la segmentación basada en contornos estáticos, es propensa a fallos [10]. Los contornos activos, en cambio, pueden ser muy eficaces para extraer la silueta del peatón, pero requieren de una buena inicialización. La aplicación de visión estéreo al campo de la detección de personas es muy reciente [14]. Permite llevar a cabo un análisis de oclusiones, es robusto ante cambios de luz, detecta tanto objetos estáticos como dinámicos, no exige un fondo estático y permite obtener medidas de las distancias a las que están los objetos [4]. Sin embargo, de esta segmentación habitualmente no se obtiene un contorno muy preciso.

El vehículo UTA [10] realiza una clasificación de los candidatos basada en la textura y la forma. Sus resultados dependen de una correcta segmentación del contorno. El vehículo ARGO selecciona aquellos objetos más afines a los rasgos humanos [4]. Para el NAVLAB se validan los objetos segmentados mediante técnicas estéreo, en función de la forma.

Los errores de estos sistemas son generados por personas próximas a la cámara [6] [4], patrones que no están contenidos en el modelo [10] [4], contornos parecidos a una persona [21], grupos de personas o zonas con una simetría vertical alta, entre otras causas.

En la actualidad se está desarrollando la fase de detección de objetos, mediante la combinación de distintas técnicas (Figura 8). La detección inicial tiene lugar aplicando visión estéreo [16]. Para tratar de mejorar la precisión de los contornos segmentados, se emplean modelos deformables definidos mediante B-Splines. Se pretende

resolver el problema de la inicialización de dichos modelos, posicionándolos en aquellas regiones de interés del mapa de disparidad. La búsqueda se realiza a distintos niveles, aplicando técnicas de desdibujado multiescala. Para la fase de reconocimiento se va a realizar un seguimiento de regiones, que permita resolver oclusiones y posibles variaciones de apariencia de un objeto debido a sombras o a cambios de iluminación. Además, a diferencia de los sistemas existentes, se realizará un tratamiento de los falsos positivos, verificando su existencia mediante la integración temporal. Hay que destacar que no se va a aplicar restricciones al movimiento ni a la apariencia de los peatones.



Fig. 8. Detección de peatones

6 Detección de los límites de la carretera

La detección de carreteras con una cámara en movimiento y en ambientes exteriores se enfrenta a dos grandes dificultades. Por una parte, los cambios bruscos de iluminación, así como la presencia de suciedad, sombras, brillos, reflejos, grietas, parches de asfalto y otros obstáculos en la carretera dificultan en gran medida el tratamiento e interpretación de la imagen. Por otra parte, el sistema debe funcionar en tiempo real y esto limita la complejidad que se le puede dar al tratamiento. Hasta ahora, la mayor parte del esfuerzo de investigación en este campo se ha dedicado a la conducción automática, para la cual es suficiente con estimar la posición y orientación del vehículo dentro del carril. Sin embargo, un sistema de apoyo al conductor requiere de unas habilidades perceptivas capaces de interpretar el entorno para predecir maniobras y situaciones de alto riesgo con antelación [22]. El requisito de tiempo real exige técnicas o suposiciones que facilitan la detección y aceleran el proceso. Las más utilizadas son: el análisis de regiones específicas, hacer suposiciones sobre el mundo (por ejemplo, que la carretera es plana o tiene un ancho constante), técnicas de optimización, o estrategias de multirresolución.

Generalmente el procesado se compone de una primera etapa de extracción de características de la imagen propias de la carretera, y una segunda etapa de ajuste de un modelo. Existen sistemas que simplifican este esquema pero éstos están orientados exclusivamente a la conducción automática, por lo que no serán tratados aquí.

En la etapa de extracción de características hay dos grandes enfoques, según la extracción se base en regiones o en bordes. Los primeros se emplean en carreteras no marcadas o no pavimentadas. No se describirán al no ser el caso del presente estudio. Las técnicas basadas en la detección de bordes buscan los límites de la calzada o las

marcas viales. Después de la detección hay un posterior agrupamiento de los píxeles marcados en estructuras de más alto nivel, es decir, líneas o marcas viales. La gran mayoría de sistemas utilizan técnicas basadas en el gradiente, aunque se han hecho esfuerzos para buscar otras técnicas de menor coste computacional como en [12] mediante segmentación basada en histograma o en [19] trabajando en el dominio de la frecuencia.

Las técnicas de agrupamiento suelen estar basadas en reglas geométricas [12] [2], en restricciones impuestas a los parámetros de un modelo [25], o en lógica borrosa [16]. En general, estas técnicas fallan cuando en la imagen hay presentes muchos bordes no pertenecientes a carretera ya que resulta difícil distinguir los que pertenecen a marca vial del resto. Asimismo, hay que tener en cuenta para la etapa de modelado de la carretera que las líneas pueden ser ocluidas por otros vehículos, obras, etc.

Una vez extraídas las características de la imagen, en la etapa de modelado se pretende obtener los parámetros del modelo deformable que se ajusten a las observaciones. El modelo debe dar una representación precisa de la carretera, ser robusto frente a oclusiones, calidad de las marcas y condiciones ambientales, ser eficiente computacionalmente, y poder colaborar con otros módulos.

Los bordes del carril se suelen modelar como líneas rectas, arcos de circunferencia sobre suelo plano [19], clotoides, splines [25] o snakes [27]. Estos modelos se pueden ajustar sobre el plano de imagen o sobre el de la carretera. Conociendo la altura e inclinación de la cámara y suponiendo que la carretera es plana [24] [2] se puede obtener un plano de la carretera a vista de pájaro, donde el ajuste del modelo es más sencillo.

El ruido presente en los entornos exteriores se suele tratar mediante el tratamiento temporal validando las observaciones comparándolas con las previas [2], o mediante un filtrado temporal [12].

Algunos autores hacen además una reconstrucción 3D de la escena, reconstruyen la carretera a base de agrupar elementos básicos (líneas o círculos) que poseen entre sí ciertas relaciones geométricas. Otros enfoques asumen un modelo paramétrico de la carretera, con etapas de inicialización y seguimiento que estiman sus parámetros en función de las características de la imagen actual [19]. Otro enfoque muy utilizado es el propuesto por [12], el cual emplea un modelo de variables de estado que incluye, además de los parámetros geométricos de la carretera y de calibración de la cámara, las ecuaciones diferenciales que relacionan movimiento del vehículo con desplazamientos espaciales. Se utiliza el filtro de Kalman para estimar las variables de estado. Este modelo a pesar de ser muy completo, es bastante sensible al ruido y requiere una costosa etapa de inicialización.

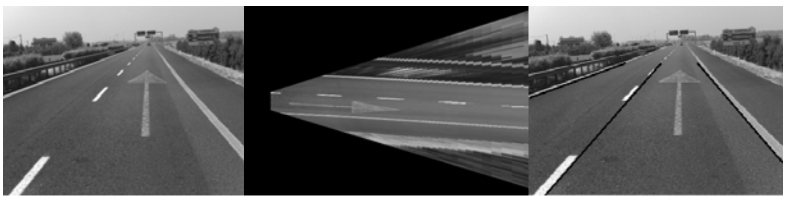


Fig. 9. Detección de las líneas de la carretera

En general, estos enfoques aportan herramientas potentes para el análisis de la carretera, pero tienen aún varios inconvenientes. Es difícil elegir y mantener el modelo apropiado, es ineficiente ajustar modelos complejos, y presentan alta complejidad computacional.

En la figura 9 se muestra la detección de los bordes del carril mediante la técnica de la perspectiva inversa y la transformada de Hough [5].

7 Conclusiones

En la actualidad se está desarrollando un sistema de asistencia a la conducción basado en visión por computador que comprende cuatro módulos interconectados: detección y clasificación de señales de tráfico, detección de vehículos, peatones y límites de la carretera. Todos ellos una vez implementados están siendo probados en la plataforma de investigación IvvI. El sistema sensorial embarcado en el vehículo está formado por una cámara color y un sistema estéreo B&W. Los sistemas de detección y clasificación de señales de tráfico y detección de los límites de la carretera ya se encuentran instalados en la plataforma y procesan la información en tiempo real, ofreciendo al conductor información de posibles peligros, maniobras incorrectas, etc.

La posición y velocidad del vehículo es proporcionada a los diferentes subsistemas por un GPS conectado a través de un enlace bluetooth con una PDA, que trasmite dicha información a los ordenadores mediante una conexión wifi.

Junto con el desarrollo de los módulos anteriormente mencionados se está trabajando en el desarrollo de un sistema de autocalibración para el sistema sensorial, pensando en su posible implantación en cualquier vehículo comercial.

Agradecimientos

Este trabajo ha sido subvencionado en parte por la CICYT a través del proyecto ASISTENTUR (TRA2004-07441-C03-01)..

Referencias

1. M. Betke, E. Haritaoglu, L.S: Davis (2000) Real-time multiple vehicle detection and tracking from a moving vehicle. Machine Vision and Applications 12, 69-83.
2. A. Broggi, M. Bertozzi, A. Fascioli, G. Conte, (1999) Automatic Vehicle Guidance: The Experience of the ARGO Autonomous Vehicle, World Scientific.
3. A. Broggi, M. Bertozzi, A. Fascioli, C. Guarino, A. Piazzi (2000) Visual perception of obstacles and vehicles for platooning. IEEE Transactions on Intelligent Transportation Systems 1 (3) 164-176.
4. Broggi, A.; Bertozzi, M.; Fascioli, A.; Sechi, M. (2000) Shape-based pedestrian detection. IEEE Intelligent Vehicles Symposium.
5. J.M. Collado, C. Hilario, A. De la Escalera, J.M. Armingol (2005) Detection and Clasification of Roas Lanes with a Frequency Analysis. IEEE Intelligent Vehicle Symposium.
6. Curio, C.; Edelbrunner, J.; Kalinke, T; Tzomakas, C.; Seelen, C. (2000) Walking pedestrian recognition. IEEE Transactions on Intelligent Transportation Systems 1 (3), 155-163.

7. A. de la Escalera, L. Moreno, M.A. Salichs, J. Mª. Armingol (1997) Road Traffic Sign Detection and Classification, IEEE Trans. on Industrial Electronics 44 (6) 848- 859
8. A. de la Escalera, J. M. Armingol, M. Mata (2003) Traffic Sign Recognition and Analysis for Intelligent Vehicles. Image and Vision Computing 11 (3) 247-258.
9. JM. Ferryman, SJ. Maybank, AD. Worrall (2000) Visual surveillance for moving vehicles. International Journal of Computer-Vision 37 (2) 187-97.
10. Gavrila, D.M. (2000) Pedestrian detection from a moving vehicle. Proc.of the European Conf. on Computer Vision.
11. D.M. Gavrila, V. Philomin (1999) Real-time object detection using distance transforms IEEE International Conference on Computer Vision.
12. J. P. Gonzalez, U. Ozguner, (2000) Lane detection using histogram-based segmentation and decision trees, IEEE Intelligent Transportation Systems.
13. U. Handmann, T. Kalinke, C. Tzomakas, M. Werner, C. Goerick, and W. von Seelen (2000) An image processing system for driver assistance. Image and Vision Computing 18, 367-376.
14. Heisele, B.; Wöhler, C. (1998) Motion-based recognition of pedestrians. Intl. Conf. on Pattern Recognition.
15. C. Hilario, J.M. Collado, J.M. Armingol, A. De la Escalera (2005) Pyramidal Image Analysis for Vehicle Detection. IEEE Intelligent Vehicles Symposium.
16. C. Hilario, J.M. Collado, J.M. Armingol, A. De la Escalera (2005) Pedestrian Detection for Intelligent Vehicles based on Active Contour Models and Stereo Vision. 10th International Workshop on Computer Aided Systems Theory.
17. C. Hilario, J. M. Collado, J. M. Armingol, A. de la Escalera (2004) Driver Assistance System Based on Computer Vision for Vehicle Detection. 5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles.
18. T. Kato, Y. Ninomiya, I. Masaki (2002) Preceding vehicle recognition based on learning from sample images. IEEE Transactions on Intelligent Transportation Systems 3 (4) 252-260.
19. C. Kreucher, S. Lakshmanan, (1999) LANA: A Lane Extraction Algorithm that uses Frequency, IEEE Transactions on Robotics and Automation, 15 (2), 343-350.
20. ND Matthews, PE An, JM Roberts, CJ Harris (1998) A neurofuzzy approach to future intelligent driver support systems. Journal of Automobile Engineering 212, 43-58
21. Papageorgiou, C.; Evgeniou, T.; Poggio, T. (1998) A trainable pedestrian detection system. IEEE Intelligent Vehicles Symposium.
22. B. Southall and C. Taylor (2001) Stochastic road shape estimation. 8th IEEE International Conference on Computer Vision.
23. Z. Sun, G. Bebis, and R. Miller (2004) On-Road Vehicle Detection Using Optical Sensors: A Review. IEEE Intelligent Transportation Systems Conference.
24. L. Vlacic, (2001) Intelligent Vehicle Technologies: Theory And Applications" Butterworth-Heinemann.
25. Y. Wang; D. Shen; E. K. Teoh, (2000) Lane detection using spline model, Pattern Recognition Letters, 21 (8), 677-689.
26. C. Wöhler, J.K. Anlauf (2001) Real-time object recognition on image sequences with the adaptable time delay neural network algorithm – applications for autonomous vehicles. Image and Vision Computing 19, 593-618.
27. A. L. Yuille, J. M. Coughlan, (2000) Fundamental limits of Bayesian inference: order parameters and phase transitions for road tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (2), 160-173.
28. Y. Zu, D. Comaniciu, M. Pellkofer, and and T. Koehler, Passing (2004) Vehicle Detection from Dynamic Background Using Robust information Fusion. IEEE International Intelligent Transportation Systems Conference.

Localización basada en lógica difusa y filtros de Kalman para robots con patas

Francisco Martín, Vicente Matellán, Pablo Barrera y Jose María Cañas

Grupo de Robótica, Universidad Rey Juan Carlos,
C/ Tulipán s/n 28933 Móstoles (Madrid), España.
{fmartin,vmo,barrera,jmplaza}@gsyc.escet.urjc.es

Resumen En la liga de 4 patas de la Robocup, equipos de 4 robots AIBO autónomos se enfrentan entre sí en partidos de fútbol. El objetivo de esta competición es presentar un entorno desafiante en que se han de resolver varios problemas relacionados con la robótica. En particular, la localización. Cada uno de los jugadores ha de estar localizado durante los partidos para que su comportamiento sea coherente. En este artículo proponemos un método de localización que combina técnicas de lógica difusa y filtros de Kalman para conseguir una localización más robusta, fiable y ligera.
Palabras Clave: Robótica móvil, fútbol robótico, localización, fuzzy, kalman

1. Introducción

Una de las habilidades más básicas que ha de ser capaz de llevar a cabo un robot móvil es la capacidad de auto-localización [4]. Esta capacidad se puede definir como la habilidad de un robot de determinar su posición en el mundo usando sus propios sensores. Las técnicas que se usan para resolver este problema varían enormemente dependiendo de los sensores disponibles en cada tipo de robot. En muchos trabajos previos, por ejemplo [9][8][6][5][2], se han propuesto soluciones a la localización de robots móviles con ruedas equipados de sensores de sonar y láser. La diferencia principal de estos trabajos con nuestra propuesta es que este tipo de robots dispone de información odométrica muy precisa y una abundante información del entorno de 360º. Estas soluciones pueden no ser aplicables en otros robots cuyos sensores sean diferentes.

En este trabajo nos centramos en el robot con patas AIBO (parte derecha de la figura 1). Este robot tiene como sensor principal una cámara situada en su cabeza. Las imágenes que se obtienen de la cámara han de ser procesadas para obtener información de ella. Otra característica de este robot es la de tener como principales actuadores de locomoción cuatro patas, lo que hace difícil obtener una información odométrica precisa.

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 310-321, 2006.

Este modelo de robot se usa en la competición de la RoboCup[1] , en la categoría de cuatro patas[2] (parte izquierda de la figura 1). La competición RoboCup es una iniciativa de investigación y educación en el ámbito internacional, que pretende fomentar el campo de la inteligencia artificial y de la robótica a base de proporcionar a los investigadores un problema estándar, donde puedan probar sus trabajos y evaluarlos.

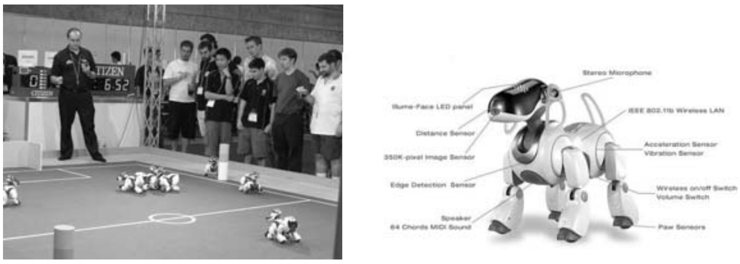


Figura 1. El TeamChaos en la RoboCup 2006 (Osaka, Japón). A la derecha el robot AIBO ERS-7

En nuestro entorno es importante que los robots estén localizados en todo momento para que su comportamiento sea coherente. Los robots no deben de salirse de los límites del campo y deben colocarse, por ejemplo, en una posición inicial conocida al iniciar el partido y después de cada gol. Así mismo, para saber hacia dónde han de dirigir la pelota y poder generar una estrategia común entre los miembros de un equipo, los robots han de estar localizados. Los métodos usados por cada uno de los equipos son muy variados, quedando reflejados en la tabla 1 los usados en la edición del 2004, última para la que está publicada la documentación de todos los equipos. En esta tabla puede apreciarse como la mayor parte de los equipos usan métodos basados en Filtros de Partículas (Localización de Monte Carlo). Varios equipos utilizan también el Filtro de Kalman Extendido, algunos de ellos combinados con Monte Carlo. Otros simplemente utilizan triangulación. Nuestro equipo[3] hasta la fecha ha usado un método de lógica difusa [1] [7].

En una competición tan exigente como la RoboCup, en la que se tiene que realizar una toma de decisiones en tiempo real, el tiempo de proceso de cada uno de sus módulos es crítico. Si se pretenden realizar comportamientos complejos de alto nivel, estrategias sobre todo, el procesado de tareas de bajo nivel, como filtrado de imágenes o localización, han de consumir el mínimo tiempo de proceso posible. El método de localización de lógica difusa, que hemos estado usando hasta ahora, consume una gran cantidad de recursos del robot si se desea una precisión de localización aceptable. Esto ha motivado que se decidiera estudiar

[1] http://www.robocup.org
[2] http://www.tzi.de/4legged/bin/view/Website/WebHome
[3] http://veo.dat.escet.urjc.es/dipta/teamchaos.html

Método	N° equipos
Localización de Monte Carlo	10
Filtro Extendido de Kalman	2
Monte Carlo + Filtro Extendido de Kalman	2
Triangulación	2

Cuadro 1. Métodos de localización utilizados en la edición de la RoboCup 2004

la combinación de este método con un Filtro de Kalman, que es un estimador de estados óptimo que consume pocos recursos computacionales. Este artículo mostrará el método usado y los resultados obtenidos, demostrando que es posible obtener una localización fiable y robusta, que es aportada por el método de lógica difusa, y una condiciones aceptables de tiempo de proceso, aportadas por el Filtro de Kalman.

La sección 2 mostrará el entorno en el que el robot se ha de localizar. En la sección 3 mostraremos el método de lógica difusa empleado en hasta ahora. En la sección 4 expondremos el método de Filtro de Kalman diseñado para este entorno, para explicar en la sección 5 cómo combinados los dos anteriores. En la sección 6 pondremos a prueba la aproximación implementada, que será discutida finalmente en la sección 7, en la que abordaremos también los posibles trabajos futuros.

2. Entorno

El entorno donde debe localizarse el robot está diseñado para que el robot disponga de una serie de marcas visuales que el robot puede usar para localizarse. Las dimensiones del campo y las posiciones de las marcas visuales son conocidas *a priori*, como puede observarse en la figura 2. Estas marcas visuales son las dos porterías y cuatro balizas de colores situadas en las bandas. Nada impide que uses otras marcas del campo, como son las líneas, para localizarte en él. Las condiciones de luz son también controladas para que sean constantes en todo momento. A pesar de estas facilidades, los algoritmos desarrollados para la localización deben tener en cuenta que las marcas pueden ser tapadas por otros robots, y que la información que se extrae de las imágenes puede ser errónea, debido al continuo movimiento de la cámara, que se encuentra situada en la cabeza del robot, las colisiones entre ellos, etc.

3. Método de localización usando lógica difusa

El objetivo de desarrollar este método de localización fue la de dotar al robot de una forma robusta de representar la incertidumbre de su posición. También debía de poder recuperarse de situaciones en las que el robot era desplazado de un lugar a otro del campo. Esta situación se puede producir cuando el robot es penalizado por el árbitro y retirado unos minutos del terreno de juego, o cuando

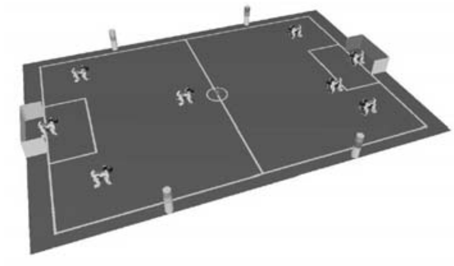


Figura 2. Campo de juego

es empujado por otros robots. Su descripción completa puede encontrarse en [1] y [3]. A continuación resumimos este trabajo previo como base de la mejora propuesta en este artículo.

El campo de juego se divide en una cuadrícula G_t tal que $G_t(x, y)$ representa la probabilidad, en $[0, 1]$, de que el robot se encuentre en la posición (x, y). Cada una de las posiciones de esta cuadrícula es una celda de dimensión configurable. Cada una de las celdas, que definiremos en adelante como *fcell*, contiene información sobre la probabilidad de que el robot esté en esa celda, e información sobre cual es el rango de orientaciones más probables para el robot, es decir, se trata realmente de una cuadrícula de $2\frac{1}{2}$D

Esta información se representa por medio de un trapezoide difuso. En la figura 3 podemos ver este trapezoide, definido por la tupla

$$< \theta, \Delta, \alpha, h, b >$$

Intuitivamente, si h es bajo, la probabilidad de estar en esta celda es baja. Si h es alto, es muy probable que el robot esté en esta posición. Si el trapezoide es ancho, existe gran incertidumbre sobre la orientación del robot. Si el trapezoide es estrecho, o tiene incluso forma de triángulo (porque Δ es prácticamente nulo), la incertidumbre de orientación es tan baja que podemos afirmar que es θ.

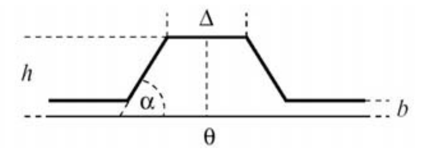


Figura 3. *fcell* representando el ángulo θ

El proceso de localización usando en este método es iterativo, teniendo cada ciclo un paso de predicción y otro de actualización. La fase de predicción se realiza cada vez que se realiza un movimiento difuminando la probabilidad en la dirección del movimiento. En la fase de actualización se incorpora la información

visual. Cada observación de una marca visual se compone de una distancia y un ángulo a cada una de las visibles. Para codificar la información de una marca visual conocida en el instante t, construimos la distribución de probabilidad $S_t(\cdot|r)$, tal que $S_t(x, y|r)$ es la posibilidad de que el robot se encuentre en la posición (x, y), siendo la distancia a una marca visual determinada r.

Un ejemplo de la aplicación secuencial de estas dos operaciones para la localización de un robot se puede observar en la figura 4. El robot parte de un estado de total incertidumbre, y mediante la información odométrica y la información de la posición relativa de la portería y de la baliza superior derecha, termina localizándose correctamente.

El proceso anterior nos proporciona la probabilidad de estar en una posición de la cuadrícula, pero no nos aporta información angular. Podría ser natural considerar un cubo 3D $G_t(x, y, \theta)$, pero el tiempo de computación de todas las posibles posiciones del robot haría que el algoritmo no fuera abordable computacionalmente. En lugar de eso se mantiene una cuadrícula 2D de *fcell*, que es una forma compacta de representar información angular. Con el procedimiento descrito anteriormente obtenemos la componente h de esta *fcell*, pero aún es necesario describir como obtener el resto de las componentes cuando se produce una observación. Esta operación es muy ligera y calcula las nuevas probabilidades de orientación a partir de una estimación anterior y una nueva observación.

4. Método de localización usando filtro de Kalman

Para implementar un Filtro de Kalman hemos definido la posición de un robot como el vector de estado $s \in \Re^3$, que se compone de :

$$\mathbf{s} = \begin{pmatrix} x_{robot} \; y_{robot} \; \theta_{robot} \end{pmatrix} \tag{1}$$

El proceso de evolución de esta estimación de la posición del robot estará guiado por dos funciones no lineales, f y h (ésta última será definida en la sección 4.2 para mayor claridad). La primera, f (obtenida a partir de la figura 4, relaciona el estado anterior s_{t-1}, la odometría u_{t-1} y un ruido en el proceso gausiano w_{t-1} con el estado actual s_t:

$$s_t = f(s_{t-1}, u_{t-1}, w_{t-1}) \tag{2}$$

$$f_1 = x_t^- = x_{t-1}^- + (u_{t-1}^x + w_{t-1}^x) cos\theta_{t-1} - (u_{t-1}^y + w_{t-1}^y) sen\theta_{t-1} \tag{3}$$

$$f_2 = y_t^- = y_{t-1}^- + (u_{t-1}^x + w_{t-1}^x) sen\theta_{t-1} + (u_{t-1}^y + w_{t-1}^y) cos\theta_{t-1} \tag{4}$$

$$f_3 = \theta_t^- = \theta_{t-1}^- + u_{t-1}^\theta + w_{t-1}^\theta \tag{5}$$

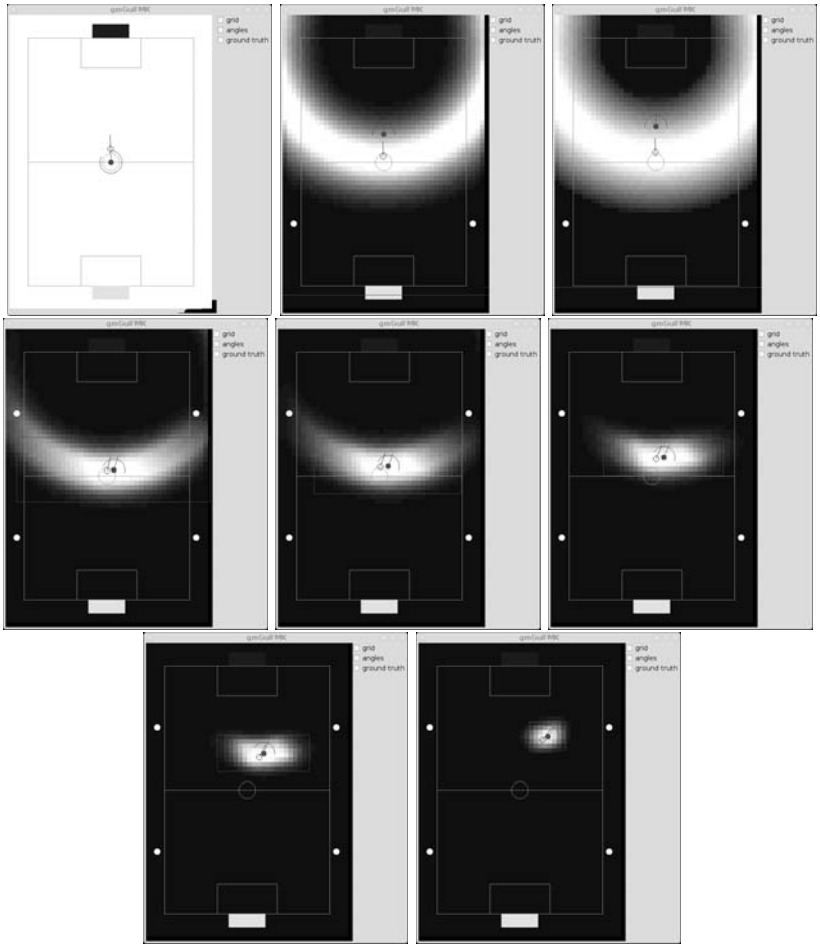


Figura 4. Proceso de localización del robot mediante grid borroso. Se parte de una situación de total ignorancia (arriba izquierda). Con la información sensorial de la portería y de la baliza izquierda el robot consigue afinar su posición para conseguir una localización fiable (derecha abajo) .

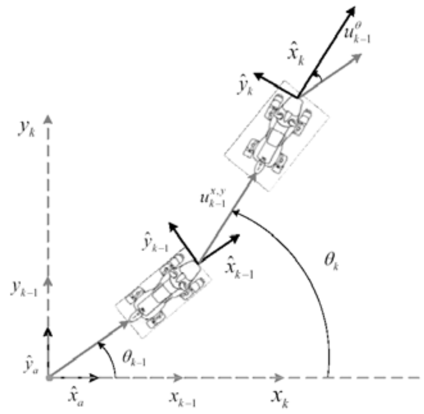


Figura 5. Modelo de movimiento basado en la odometría

4.1. Fase de predicción

En esta fase calcularemos $s_t^- \in \Re^3$ que es la estimación de la posición a priori, es decir, la posición en la que se encontrará el robot según la información de la odometría y aplicando el modelo de movimiento de la figura 4 en el proceso. También calcularemos $P_t^- \in \Re^{3x3}$, que es la covarianza del error del sistema a priori. Esta matriz será la que almacene el error general del sistema. Si tuviéramos certeza absoluta de la posición inicial $P_0 = 0$, pero este no es el caso. Arbitrariamente elegimos una matriz no nula, $P_0 = I_3$, ya que esta matriz identidad es capaz de converger rápidamente a valores correctos.

$$s_t^- = f(s_{t-1}, u_{t-1}, 0) \tag{6}$$

$$P_t^- = A_t P_{t-1} A_t^T + W_t Q_{t-1} W_t^T \tag{7}$$

A continuación vamos a describir los principales elementos que intervienen en la fase de predicción y el significado de sus valores:

Q_t es la covarianza del error en la fase de predicción viene definido por $Q_t \in \Re^{3x3}$ ($Q_t = E[w_k w_k^T]$), y los valores que contiene son experimentales. Indica el error que se puede producir al aplicar la odometría.

$$Q_t = \begin{pmatrix} 0{,}1|u_{t-1}^x| & 0 & 0 \\ 0 & 0{,}1|u_{t-1}^y| & 0 \\ 0 & 0 & 0{,}1|u_{t-1}^\theta| + 0{,}001\sqrt{(u_{t-1}^x)^2 + (u_{t-1}^y)^2} \end{pmatrix} \tag{8}$$

A_t es el jacobiano de f con respecto a s. De las ecuaciones 3, 4 y 5 obtenemos:

$$A_t = \frac{\partial f}{\partial s} = \begin{pmatrix} 1 & 0 & -u_{t-1}^y cos\theta_{t-1} - u_{t-1}^x sen\theta_{t-1} \\ 0 & 1 & u_{t-1}^x cos\theta_{t-1} - u_{t-1}^y sen\theta_{t-1} \\ 0 & 0 & 1 \end{pmatrix} \tag{9}$$

W_t es el jacobiano de f con respecto al error de odometría w. Al igual que en 9, de las ecuaciones 3, 4 y 5 obtenemos:

$$W_t = \frac{\partial f}{\partial w} = \begin{pmatrix} cos\theta_{t-1} & -sen\theta_{t-1} & 0 \\ sen\theta_{t-1} & cos\theta_{t-1} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{10}$$

4.2. Fase de corrección

En la fase de corrección, la posición del robot se corrige con la información sensorial que percibe. Esta información sensorial percibida, $\hat{z}_t \in \Re^{2m}$, tiene dimensión variable, pues depende del número de elementos m del que tiene información. La medida para cada elemento está formada por un ángulo (α_i) y una distancia (ρ_i) a dicho elemento. De esta manera, tenemos:

$$\hat{\mathbf{z}}_\mathbf{t} = \begin{pmatrix} \rho_1 \; \alpha_1 \; \rho_2 \; \alpha_2 \cdots \rho_m \; \alpha_m \end{pmatrix}^T \tag{11}$$

En esta fase hallaremos la posición del robot una vez corregida con las medidas s_t y la matriz de covarianza P_t que almacena el error del sistema:

$$s_t = s_t^- + K_t(\hat{z}_t - h(s_t^-, 0)) \tag{12}$$

$$P_t = (I - K_t H_t) P_t^- \tag{13}$$

$$K_t = P_t^- H_t^T (H_t P_t^- H_t^T + V_t R_t V_t^T)^{-1} \tag{14}$$

La función $h(s_t^-, 0)$ relaciona la posición s_t^- con la medida teórica z_t que debería obtenerse en esa posición. Usamos un modelo geométrico para calcularlo:

$$z_t = h(s_t^-, 0) \tag{15}$$

$$\begin{pmatrix} \rho_1 \\ \alpha_1 \\ \vdots \\ \rho_m \\ \alpha_m \end{pmatrix} = \begin{pmatrix} h_1(s_t^-, 0) \\ h_2(s_t^-, 0) \\ \vdots \\ h_{2m-1}(s_t^-, 0) \\ h_{2m}(s_t^-, 0) \end{pmatrix} = \begin{pmatrix} distancia(s_t^-, lm_1) \\ angulo(s_t^-, lm_1) \\ \vdots \\ distancia(s_t^-, lm_m) \\ angulo(s_t^-, lm_m) \end{pmatrix} \tag{16}$$

A continuación vamos a describir los principales elementos que intervienen en la fase de corrección y el significado de los valores aplicados:

R_t Las medidas tienen asociadas una matriz, $R_t \in \Re^{2mx2m}$, $R_t = E[v_k v_k^T]$, que representa el error en la medida. Esta matriz es una matriz diagonal cuyos valores son empíricos. En el caso de las distancias, el valor depende del tipo de elemento y de la distancia a él, y representa el error que puede darse en esas condiciones. En caso del ángulo, el error es muy pequeño (2°) si el elemento está completo en la imagen, y mayor si no está completo (18°).

V_t La matriz $V_t \in \Re^{2mx2m}$ es la matriz del jacobiano de h con respecto a v, que es la matriz identidad I^{2m} en nuestro caso.

H_t la matriz $H_t \in \Re^{2mx3}$ es la matriz del jacobiano de h con respecto a s,

$$H_t = \frac{\partial h_t}{\partial s^-} = \begin{pmatrix} \frac{\partial distancia(s_t^-, lm_1)}{\partial x} & \frac{\partial distancia(s_t^-, lm_1)}{\partial y} & \frac{\partial distancia(s_t^-, lm_1)}{\partial \theta} \\ \frac{\partial angulo(s_t^-, lm_1)}{\partial x} & \frac{\partial angulo(s_t^-, lm_1)}{\partial y} & \frac{\partial angulo(s_t^-, lm_1)}{\partial \theta} \\ \vdots & \vdots & \vdots \\ \frac{\partial distancia(s_t^-, lm_{2m-1})}{\partial x} & \frac{\partial distancia(s_t^-, lm_{2m-1})}{\partial y} & \frac{\partial distancia(s_t^-, lm_{2m-1})}{\partial \theta} \\ \frac{\partial angulo(s_t^-, lm_{2m})}{\partial x} & \frac{\partial angulo(s_t^-, lm_{2m})}{\partial y} & \frac{\partial angulo(s_t^-, lm_{2m})}{\partial \theta} \end{pmatrix} \quad (17)$$

$$H_{[i,j],t} = \frac{\partial h_{t,i}}{\partial s_j} = \frac{h_i(s_{j,t}^- + \Delta, 0) - h_i(s_{j,t}^- - \Delta, 0)}{2\Delta} \quad (18)$$

5. Combinación de lógica difusa y filtro de Kalman

La combinación de los dos métodos de localización combina las mejores cualidades de cada uno. Por una parte, el método de localización basado en lógica difusa permite una localización robusta, capaz de recuperarse de situaciones de completa incertidumbre.

Este método, aunque use una manera compacta de representar la información angular de cada casilla para hacerlo más ligero, consume muchos recursos computacionales. Más aún cuando para la edición del 2005 de la RoboCup se decidió aumentar considerablemente la resolución[4]. Una posible solución para que el método fuera usable sería aumentar el tamaño de cada cuadrícula en la que se divide el campo, pero esto daba lugar a una significativa pérdida de precisión en la estimación de la posición del robot.

El método de localización basado en el Filtro de Kalman es, por contra, computacionalmente ligero. El problema de este método es que no es capaz de localizarse partiendo de un escenario de total incertidumbre, y no es capaz de recuperarse de situaciones de alto error en la estimación o de cambio manual de la posición del robot en el terreno de juego.

La estrategia propuesta consiste en intercambiar el método de localización alternativamente. Al iniciarse el robot, éste usa el método de localización basado en lógica difusa hasta que la calidad de esta localización es suficientemente aceptable para asegurar que el robot está perfectamente localizado. En ese momento el robot cambia su algoritmo de localización al Filtro de Kalman. Si en algún momento la matriz P del Filtro de Kalman refleja una incertidumbre excesiva para considerar al robot localizado, el algoritmo de localización del robot cambia al método de lógica difusa. El proceso se repite alternativamente dependiendo de la calidad de las estimaciones.

[4] http://www.tzi.de/4legged/pub/Website/Downloads/Rules2005.pdf

6. Experimentos

Los experimentos llevados a cabo en este proyecto se centran en validar la robustez del método del Filtro de Kalman. Los resultados experimentales del método de lógica difusa pueden encontrarse en [1][3].

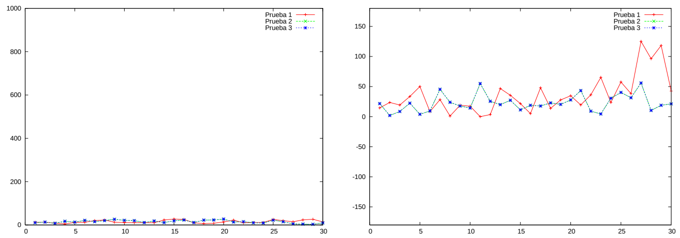


Figura 6. Error en la estimación de la posición (x, y) y error absoluto en la estimación de la orientación θ en el experimento 1

Los experimentos llevados a cabo para validar el método del Filtro de Kalman se llevaron a cabo en un simulador que permite configurar los valores máximos de un ruido que es generado aleatoriamente, tanto para la información odométrica, como para las medidas de distancia y ángulo con respecto a cada una de las marcas visuales.

Las figuras 6,7 y 8 muestran los resultados de tres experimentos diferentes donde el robot ha realizado un desplazamiento por el terreno de juego. Las figuras muestran el error en la estimación de la posición (x, y) y en la estimación de la orientación θ con valores de ruido similares a los que realmente existen en la realidad.

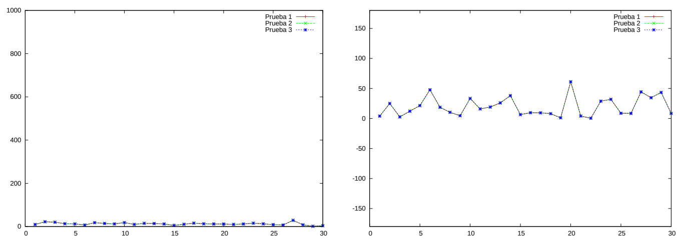


Figura 7. Error en la estimación de la posición (x, y) y error absoluto en la estimación de la orientación θ en el experimento 2

En estos experimentos se muestra como el error en estimación de la posición (x, y) no supera en ningún momento los 7 cm. La estimación de orientación es

la que más error acumula, aunque nunca llega a superar los 60 grados. Además, se mantiene bajo a los largo de la mayoría del proceso.

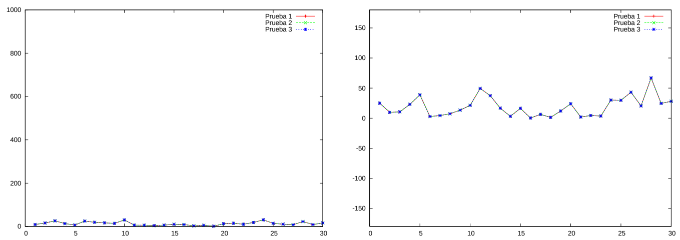


Figura 8. Error en la estimación de la posición (x, y) y error absoluto en la estimación de la orientación θ en el experimento 3

Hay que hacer constar que no se produce un cambio continuo de métodos de localización, ya que el método de lógica difusa converge rápidamente, y el Filtro de Kalman estima correctamente la posición en la mayor parte las situaciones. Generalmente las únicas situaciones en las que hay que volver a localizarse totalmente con el método de lógica difusa son al principio del funcionamiento del robot, cuando éste es “secuestrado”, o cuando hay un error grave y continuo de la información extraída de las imágenes.

7. Conclusiones y trabajo futuro

Se ha desarrollado en este trabajo un método de localización combinando la robustez y capacidad de recuperación del método de localización difuso implementado previamente, con un método que usa un Filtro de Kalman, cuyo punto fuerte es el escaso tiempo de procesamiento que necesita debido a las operaciones que ha de realizar en cada ciclo. El Filtro de Kalman desarrollado es capaz de estimar correctamente la posición del robot con los niveles de ruido presentes en el escenario donde se ha de localizar. La combinación de ambos métodos es simple. Se activa en cada momento el método adecuado dependiendo de las necesidades de localización y de la estimación actual. La mayor parte del tiempo usa el método del Filtro de Kalman, que consume pocos recursos, y el método de localización basado en lógica difusa se activa puntualmente para recuperarse rápidamente en caso de que la estimación del error aumente mucho.

El tiempo de procesamiento total del módulo de localización se ha optimizado, resultado que era necesario para la viabilidad de la totalidad del código desarrollado para el robot. Si este tiempo fuera superior, tareas como la locomoción o la coordinación del movimiento con la estimación de la pelota fallarían.

Los trabajos futuros se centran en realizar un combinación óptima de ambos algoritmos. También el desarrollo de otros métodos, como puedes ser Filtros de Partículas, para reiniciar el Filtro de Kalman en ciertas situaciones puede ser útil.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Ciencia y Tecnología, en el proyecto ACRACE: DPI2004-07993-C03-01 y la Comunidad de Madrid en el proyecto RoboCity 2030: S-0505/DPI/0176.

Referencias

1. P. Buschka, A. Saffiotti, and Z.Wasik. Fuzzy landmark-based localization for a legged robot. In *Proceedings of the International Conference on Intelligent Robots and Systems 2000*, Takamatsu, Japan, October 2000.
2. Dieter Fox, Wolfram Burgard, and Sebastian Thrun. Markov localization for mobile robots in dynamic environments. *Journal of Artificial Intelligence Research*, 11:391–427, 1999.
3. D. Herrero-Pérez, H. Martínez-Barberá, and A. Saffiotti. Fuzzy self-localization using natural features in the four-legged league. *Lecture Notes in Computer Science. Robocup 2004*, 3276, 2005.
4. J.Borenstein, B.Everett, and L.Feng. *Navigating mobile robots: Systems and techniques.* Ltd. Wesley, MA, 1996.
5. Jana Kosecká and Fayin li. Vision based topological markov localization. In *Proceedings of the 2004 IEEE International Conference on Robotics and Automation*, Barcelona (Spain), April 2004.
6. María E. López, Luis Miguel Bergasa, and M.S.Escudero. Visually augmented POMDP for indoor robot navigation. *Applied Informatics*, pages 183–187, 2003.
7. Humberto Martínez, Vicente Matellán, and Miguel Cazorla. Teamchaos technical report. Technical report, TeamChaos, 2006.
8. Dandapani Radhakrishnan and Illah Nourbakhsh. Topological localization by training a vision-based transition detector. In *Proceedings of IROS 1999*, volume 1, pages 468 – 473, October 1999.
9. Reid Simmons and Sven Koening. Probabilistic navigation in partially observable environments. In *Proceedings of the 1995 International Joint Conference on Artificial Intelligence*, pages 1080–1087, Montreal (Canada), July 1995.

Reflexiones sobre la utilización de robots autónomos en tareas de vigilancia y seguridad

José R. Álvarez Sánchez, José Mira y Félix de la Paz López

Departamento de Inteligencia Artificial UNED
{jras,jmira,delapa}@dia.uned.es

Resumen La utilización de robots móviles en tareas de vigilancia, inspección y seguridad en general va a ser uno de los campos naturales de expansión de este tipo de máquinas en los próximos años. Sin embargo, presentan algunas limitaciones. En este artículo se realiza una evaluación de las posibilidades de utilización de robots autónomos en tareas de seguridad y se expone como caso práctico el estudio de viabilidad de su aplicación en el caso concreto del proyecto AVISA.
Palabras clave: robótica autónoma; vigilancia y seguridad; visión activa

1. Tareas de vigilancia y seguridad: la aproximación propuesta en el proyecto AVISA

El propósito general de las tareas de vigilancia y seguridad excede con mucho el alcance de un proyecto concreto. Monitorizar de forma completa un escenario genérico, ser capaces de predecir y diagnosticar potenciales situaciones de riesgo y actuar de forma eficiente para evitar esas situaciones es, claramente, un objetivo a muy largo plazo cuya importancia es difícil de exagerar.

En el proyecto AVISA [6], hemos adoptado una posición más acotada en objetivos y situaciones, aunque quizás todavía excesiva, a la luz de lo aprendido en la primera fase de nuestro trabajo en el proyecto. El proyecto AVISA pretende modelar, operacionalizar e implementar un conjunto de componentes reutilizables (agentes) en la síntesis de sistemas semiautomáticos de vigilancia, en un conjunto de escenarios reales, aunque simplificados, en los que existan situaciones de movimiento de personas, vehículos y otros objetos. Se consideran inicialmente tres tareas básicas (figura 1):

1. Monitorizar el entorno usando distintas fuentes de información fijas, manipulables o móviles.
2. Diagnosticar las situaciones de interés en términos de las clases a que pertenecen las relaciones espacio-temporales entre distintos objetos marcados en una secuencia de imágenes.
3. Generar las acciones pertinentes ante situaciones de prealarma o alarma, siempre bajo la decisión final del operador humano de una central de alarmas.

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 322-331, 2006.

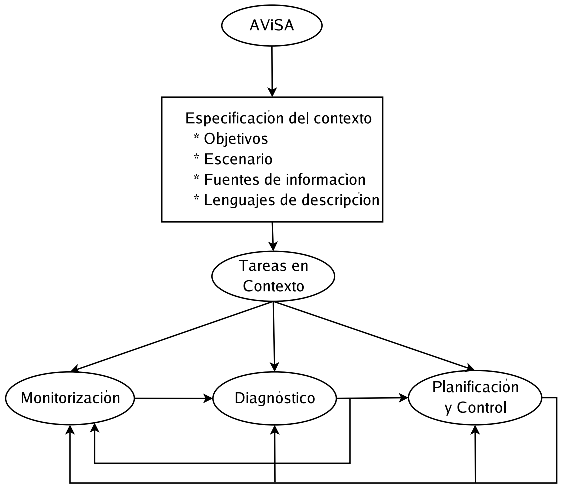


Figura 1. Subtareas básicas del proyecto AVISA[7,8].

La elección de una solución semiautomática acota los objetivos de AVISA que no pretende sustituir al operador humano sino ayudarle ante potenciales pérdidas de atención, en la síntesis de mensajes de prealarma y en el acceso a la toma de datos en entornos de difícil acceso o peligro potencial. Adicionalmente, la simplificación de los escenarios de prueba iniciales (corredores, pocos actores, repertorio limitado de situaciones etc...) también acota estos objetivos.

Finalmente, como en este trabajo sólo vamos a reflexionar sobre la utilización de robots móviles, nos vamos a referir solamente a la tarea de monitorización, dando además por supuesto que otros miembros del proyecto nos van a proporcionar la información necesaria para poner en contexto esta tarea. En particular, la especificación del escenario y los lenguajes de descripción ("blobs" o manchas, objetos, eventos y comportamientos). Es decir, sea cual fuere la naturaleza de un robot autónomo y del escenario, no vamos a considerar las cuestiones relacionadas con los procesos deliberativos que no se pueden realizar en tiempo real, asociados a la tarea de diagnóstico y planificación. Sólo nos preocupamos de los robots (a) Como agentes que son capaces de colocar un sensor y/o un efector en el sitio y en el tiempo que es necesario para ayudar a monitorizar un escenario y (b) como generadores de perspectivas de toma de datos que son necesarias o convenientes para que el operador de una central de alarmas disponga en tiempo real de la mayor cantidad posible de información relevante sobre la evolución temporal del estado del espacio que se está vigilando y de acuerdo con el propósito de su tarea de vigilancia.

El resto del trabajo está organizado de la siguiente forma. En la sección 2 mencionamos los tres tipos de información (sensores fijos, teleoperación y robots móviles). Después enumeramos una serie de tareas específicas, bajo el marco general de vigilancia, en las que es razonable incorporar robots móviles (sección 3).

En la sección 4 ofrecemos algunos datos sobre la conveniencia de utilizar robots en función de la selección entre su coste y el valor económico o estratégico de los bienes a cuya protección se va a dedicar el robot. En la sección 5 abordamos el serio problema de la integración de fuentes de información de posición controlable para obtener la visión más útil de un escenario, de acuerdo con los objetivos de la tarea de vigilancia, las limitaciones físicas del entorno y las situaciones de interés (sombras, oclusiones, zonas no accesibles, ruido, etc...). Finalmente, concluimos poniendo de manifiesto la dificultad de las tareas y alguna de las limitaciones de la robótica actual para convertirse en eficaz colaboradora del operador humano de una central de vigilancia.

2. Sensores fijos, teleoperación y robots móviles

El tipo más usual y sencillo de sensor en una instalación de seguridad es el sensor fijo, ya sea un sensor volumétrico, de barrera o una cámara. Para este tipo de sensores, la tarea de monitorización se limita a seguir la evolución temporal de sus datos, con posibilidad de focalizar sobre aquellos que se consideran dinámicamente de más interés (por ejemplo, por que hay movimiento).

Por otra parte, la propia naturaleza de las tareas de vigilancia y seguridad convierten los escenarios de aplicación en lugares arriesgados para los mecanismos de vigilancia estáticos al ser más *vulnerables por su ubicación fija* y posiblemente conocida por los intrusos (que por tanto pueden inutilizarlos o evadirse de su detección). Por estos motivos tiene sentido considerar la utilización de dispositivos de detección y visión móviles que pueden aportar información desde diversos puntos de vista y mantener posiciones cambiantes que puedan contrarrestar ataques o destrucción de los dispositivos y además dificultar la evasión de su detección.

Adicionalmente, las características de cooperación y apoyo que los sistemas de vigilancia semi-automáticos prestan a los vigilantes humanos, se puede ver reforzada o incluso imprescindible en situaciones de *peligro para la integridad de las personas*, en las cuales los vigilantes o bien no pueden entrar o acercarse para patrullar ciertos escenarios, o bien no pueden ir a comprobar una zona donde los sensores fijos han dejado de transmitir datos fiables (cambio de condiciones que impiden la detección correcta, fallo del dispositivo, accidente o ataque por intrusos).

La autonomía en robots móviles es un problema que todavía no está resuelto en cuanto a aplicaciones en tiempo real en entornos reales [9]. En muchos casos, cuando se trata de aplicaciones críticas que requieren un grado mínimo de fallos (manipulación de explosivos, sondas robóticas espaciales etc.) se recurre a la teleoperación a través del control por un operador humano. Podemos afirmar que podemos acudir a la autonomía en cuanto a tareas de navegación, exploración y modelado del medio [3], pero cuando la aplicación es de grado más fino, por ejemplo manipulación de objetos, elección de blancos para un arma o toma de muestras para un sensor específico, sigue siendo más recomendable la teleoperación.

Es evidente que, en aplicaciones en seguridad y vigilancia, lo más importante es que el robot haga bien la tarea, lo más rápido posible y sin posibilidad de error. Es por esto que, teniendo en cuenta el estado actual de la tecnología, sea más recomendable la teleoperación. Un robot completamente autónomo está desaconsejado para éste tipo de aplicaciones a día de hoy. No obstante, es interesante explorar soluciones híbridas, como el concepto de "adjustable autonomy" propuesto en [?]. Una aplicación de éste concepto usando interfaces avanzadas, que simplifican su uso por los vigilantes humanos, puede encontrarse en [5].

Por lo tanto, también en las aplicaciones de robots móviles en vigilancia debemos considerar como una opción más la posibilidad de utilizar teleoperación total o parcial. De esta forma, el robot se convierte en una herramienta controlable por el vigilante humano al que se le permite decidir la evolución temporal de las coordenadas de toma de datos y ejecución de determinadas acciones.

3. Robots móviles en tareas de vigilancia según su aplicación

Existen numerosas tareas asociadas a la vigilancia donde se puede incorporar la ayuda de un robot móvil e incluso en las que se puede plantear la sustitución total del operador humano en favor del robot. En la figura 2 podemos ver el robot Nomad-200 asociado al proyecto AVISA que usamos como prototipo. Este tipo de robot puede ser dotado con sistemas de visión artificial (visual e IR), sensores para todo tipo de gases, incluso agentes nerviosos, radares, sensores de rango, *Lidar,* etc. Hemos resumido a continuación las principales tareas que son susceptibles de realizar por un robot especializado en vigilancia.



Figura 2. Robot Nomad-200 usado en la fase inicial del proyecto AVISA.

Salvaguarda de la integridad personal. Cuando los posibles intrusos pueden usar armas o sustancias tóxicas o paralizantes contra los vigilantes humanos es

evidente que es innecesario arriesgar una vida humana. En este caso, un robot con la suficiente capacidad sensorial es la herramienta más indicada, si el propio entorno que se vigila es por sí mismo peligroso o se puede convertir en peligroso para las personas (a consecuencia de un accidente o ataque), por la producción o almacenamiento de sustancias tóxicas o contaminantes (químicas o biológicas), sistemas anti-incendios por expulsión de oxígeno, radiactividad, altos voltajes (distribución eléctrica), explosivos, sustancias inflamables etc. Los robots usados por los equipos de desactivación de explosivos son un buen ejemplo de este tipo de tareas.

Ayuda en rondas de vigilancia. Si la cantidad de personas asignadas a la vigilancia no recomienda el abandono de un centro de mando o puesto control ante una situación anómala en un área grande vigilada (p. ej.: varias naves en un polígono industrial), un robot puede desplazarse autónomamente o por teleoperación al lugar que el centro de mando estime oportuno. También puede un robot complementar las tareas de patrulla por recintos grandes, para evitar la repetición y monotonía de rutas de patrulla sistemática (rondas) a los humanos (que los hace más predecibles, y por tanto vulnerables o más fáciles de evadir). Usando robots que no se cansan y que pueden variar continuamente de forma aleatoria sus rutas modificándolas según pequeñas variaciones o anomalías detectadas por sus sensores se puede mejorar la eficacia de las rondas.

Como sensor especializado. Cuando la especificidad y carestía de la medida sensorial a realizar (por ejemplo, radar de penetración, LADAR 3D, ...) no justifica económicamente el emplazamiento de los sensores en todos los puntos del recinto a vigilar, pero es necesario realizar esa medida bajo algunas circunstancias, podemos llevar el equipo de medida en un robot a distintos lugares.

Como herramienta de test de instalaciones de vigilancia. Un robot puede tener asignada una trayectoria de test cada cierto tiempo para comprobar el estado de funcionamiento de las cámaras y sensores instalados en el recinto vigilado, haciendo saltar o activando a propósito detectores y alarmas.

Como arma táctica. Algunos ejércitos, policías y cuerpos de seguridad especializados empiezan a incorporar robots dotados de armamento táctico, fundamentalmente en tareas de asalto [10,4].

4. Robots móviles en tareas de vigilancia según su coste

En algunos casos se puede considerar la utilización de dispositivos móviles para complementar las funciones de los dispositivos fijos de detección y visión. El tipo de sistema de transporte o movimiento de los sensores debe ser adecuado a las características del entorno o escenario que se protege o vigila. La característica principal que determina el tipo de movilidad (y también parte del riesgo para los vigilantes humanos) es el *valor económico o estratégico de los bienes protegidos.*

Este valor es el que marca directa o indirectamente el coste aceptable de inversión en dispositivos de vigilancia apropiados al escenario. Amortizar un robot de coste muy alto, teniendo en cuenta sus costes de mantenimiento asociados, no es una tarea fácil.

Los tipos de movilidad pueden variar según ese coste, de acuerdo con los siguientes niveles:

1. Sistemas de movimiento de sensores (principalmente cámaras) sobre raíles o en brazos articulados que aumentan los ángulos de visión. Coste bajo.
2. Simples plataformas de transporte para acarrear sensores (especialmente cámaras) teledirigidas directamente mediante radiocontrol por los vigilantes humanos. Coste medio.
3. Pequeñas plataformas móviles sencillas (mapa fijo preinstalado) y de bajo coste pero en gran cantidad con capacidades simples cooperativas. Coste alto.
4. Plataformas autónomas de mayor coste y mayores capacidades de navegación (mapa autogenerado) y de control (órdenes de alto nivel) en las que los propios sensores de navegación del robot (sonar, láser, etc.) se pueden utilizar como detectores o fuentes de información adicional. Coste muy alto.

Para elegir el tipo de plataforma móvil apropiada para un escenario, también hay que tener en cuenta otras restricciones, como el tamaño y las características geométricas de las superficies por donde se debe mover (suelo liso, escalones, terreno rugoso, distancias a recorrer sin recargar baterías, huecos de paso, puertas y obstáculos móviles, etc.), necesidades de control y guía por humanos, necesidades de mantenimiento, autonomía, capacidades de comunicación, etc. Además, habría que considerar la alternativa de tener sistemas fijos de detección más baratos y suplirlos con algún robot para ampliar las capacidades de forma dinámica y puntual, en aquellas coordenadas y situaciones en las que se detecte alguna anomalía, con lo cual el coste o inversión total podría ser menor, aumentando a la vez la versatilidad del sistema.

Un aspecto adicional del uso de robots en la vigilancia y seguridad es la posibilidad de incorporar *sistemas de seguridad activa* como dardos o gases paralizantes contra intrusos, extintores de incendios, etc.

5. Integración de sensores

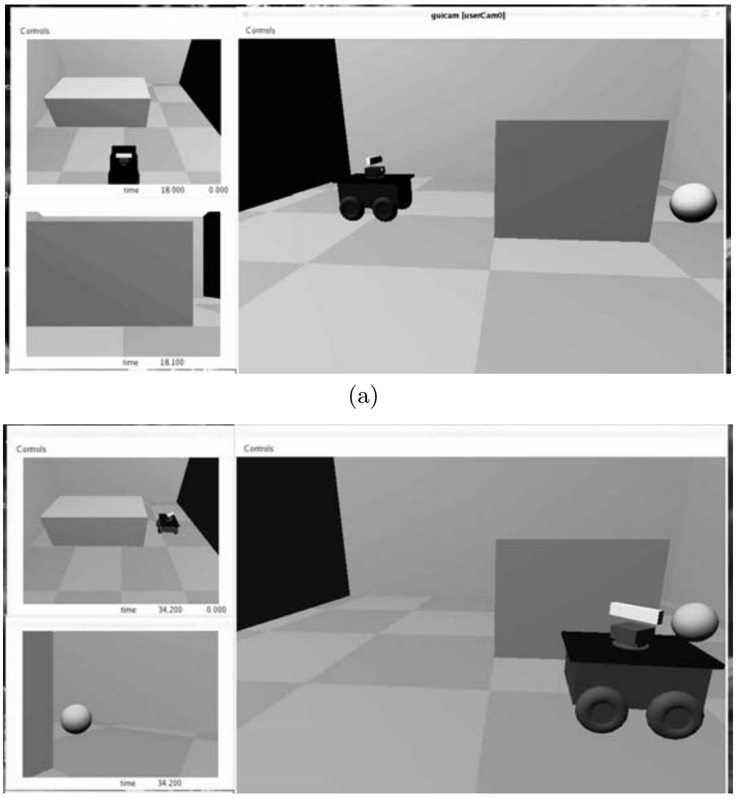
La utilización de sensores en diferentes puntos de un escenario requiere la integración de datos provenientes de múltiples sensores según su posición espacial y en relación a su secuencia temporal [2]. El uso de sensores en plataformas móviles es un caso más general que incluye las variaciones en el tiempo de las posiciones de los sensores y donde los sensores fijos son un caso particular. Los sensores más complejos y que mayor cantidad de información aportan son las cámaras. La integración de datos a bajo nivel es más sencilla con otros tipos de sensores, pero las cámaras requieren un tratamiento especial por las fuertes interrelaciones geométricas existentes en la información que aportan y por la

riqueza y variabilidad de las implicaciones semánticas que se pueden asociar a cada forma específica de integrar la información procedente de esas fuentes.

Normalmente se utilizan cámaras para detección simple de variaciones en una escena (interpretación limitada de escenas) junto con información de otros sensores para avisar a un vigilante humano de un evento o cambio de estado que puede ser de posible interés para la seguridad (ayuda al diagnóstico de eventos). En este caso, el problema genérico es la utilización de múltiples cámaras con posiciones conocidas (fijas o sobre robot situado), del cual la estereovisión es, quizás, el caso particular más simple (2 cámaras en el mismo plano separadas poca distancia y orientadas a un punto común), para el análisis/interpretación de secuencias temporales sincronizadas de proyecciones (imágenes/fotogramas) 2D cruzadas en 3D (debe haber un solape o adyacencia de esas proyecciones).

Conviene recordar sin embargo que en una tarea de vigilancia no es en general preciso disponer de una representación tridimensional completa de todos los objetos de la escena. Ocasionalmente puede ser necesario focalizar sobre un rostro e intentar obtener el máximo de información, por ejemplo para tareas de identificación. Sin embargo, en general la monitorización nos pide una representación ergonómica, con una fuerte reducción en la dimensionalidad del espacio de entradas. Es decir [7,8], una degradación de la información procedente de varias fuentes que pasa de una o varias secuencias de imágenes (píxeles en tiempo continuo) a un vector discreto, y en ocasiones booleano, de descriptores de la escena que sólo consideran el valor de ciertas propiedades espacio-temporales en determinados intervalos de muestreo. Los descriptores, su medida y el periodo de muestreo son función de la dinámica de la escena y del *tiempo propio* de los sucesos de interés. Esto supone un gran esfuerzo, previo al uso del robot, en la especificación de lo que se debe considerar como *"suceso de interés"*, porque en esta especificación está implícita la información correspondiente al diseño de la configuración de sensores y al procedimiento efectivo de integración de las distintas informaciones proporcionadas por esos sensores. Algunos de estos sucesos de potencial interés son los siguientes:

Oclusión de objetos. En muchos casos, una cámara no puede cubrir todo lo que ocurre en un escenario de vigilancia, fundamentalmente si un objeto es ocultado detrás de otro que impide su visión directa por la cámara. En estos casos, la intervención de un robot móvil dotado de una cámara puede solucionar este problema tal y como se puede ver en el ejemplo simulado de la figura 3. En este ejemplo, en una instalación de seguridad una cámara fija (ventana superior izquierda) muestra una escena en la que un objeto grande representado por un prisma rectangular oculta otro más pequeño, representado por una esfera. La escena puede observarse en su totalidad en la ventana derecha, cuya vista evidentemente no pertenece a la instalación de seguridad sino que la hemos añadido para comprender el ejemplo. El robot lleva una cámara que puede visualizarse en la ventana inferior izquierda. El vigilante humano opera el robot y la cámara para acceder a la zona no visible por la cámara fija, y descubre la esfera que estaba oculta.



(a)

(b)

Figura 3. Uso de cámaras móviles para resolver el problema de la oclusión de objetos (simulador Gazebo [1]).

Identificación y toma de muestras. A veces aparecen objetos en la escena de vigilancia que es necesario identificar, manipular o realizar pruebas sobre ellos. Normalmente esto requiere la intervención de un operario humano que normalmente tiene que dejar su puesto de vigilancia para acudir a la escena, poniendo en riesgo su integridad personal y comprometiendo durante esta intervención la seguridad del resto de escenas que controla desde su puesto. En este caso es evidente que la intervención de un robot que pueda manipular, monitorizar con sus sensores y tomar muestras está más que justificada.

Suplencia de instalación. Durante la intrusión en una instalación vigilada, puede que algunos sensores sean manipulados o inutilizados. A efectos de la central de vigilancia, esto puede manifestarse simplemente porque en un monitor aparece "nieve" y no se sabe si es un fallo técnico o un fallo provocado. En este caso podemos mandar al robot a esa escena, para averiguar si lo que está ocurriendo

es una intrusión o un fallo. En caso de fallo podemos dejar al robot allí hasta que el servicio técnico venga a reparar el equipo estropeado.

6. Conclusiones

De estas reflexiones sobre el uso de robots en tareas de vigilancia y seguridad se pueden deducir, al menos, las siguientes conclusiones básicas:

1. Ha avanzado más la componente simbólica o representacional de la inteligencia artificial (IA), y por consiguiente de la robótica, que la componente situada. Dicho de forma sencilla, sabemos más sobre cómo diseñar sistemas de IA relacionados con las tareas de percepción, decisión y planificación que sobre el desarrollo de sensores y efectores en un robot físico, tales que permitan usar esas funciones simbólicas en un entorno real, con escaleras, irregularidades etc. Estas limitaciones en movilidad son un claro inconveniente en el uso eficaz de robots en vigilancia y aconsejan seguir usando la teleoperación y el uso sobreabundante de cámaras fijas para complementar la función del robot.
2. Aún en el caso de plataformas móviles y autónomas, de coste muy alto, persisten las limitaciones de movilidad mencionadas previamente. Además, en estos casos hay que valorar la relación coste-beneficio, si bien en temas de seguridad el valor humano y estratégico (el bien protegido), no es comparable con el coste económico de la protección.
3. Para muchos trabajos asociados al uso de robots en vigilancia no es necesario disponer inicialmente de un robot porque el trabajo serio está en el estudio de la contribución de una o varias fuentes móviles de toma de datos a la construcción dinámica de la mejor representación posible del espacio tridimensional, de acuerdo con los objetivos específicos de la tarea de vigilancia. Si el problema de la integración de fuentes de información es en general complejo, esta complejidad crece al convertirse en dinámica, con fuentes móviles y con objetos cambiantes.

Si no se resuelven primero estos problemas, la presencia del robot físico puede ser meramente decorativa.

Agradecimientos

Agradecemos la financiación de este trabajo proporcionada por el Ministerio de Ciencia y Tecnología con el proyecto AVISA (Proyecto Coordinado de I+D MCyT, ref. TIN2004-07661-C02-01).

Referencias

1. Player / Stage and Gazebo Simulators. `http://playerstage.sourceforge.net/`.
2. Johann Borenstein, H.R. Everett, y L. Felng. "Where am I?". Sensors and methods for mobile robot positioning. informe técnico, University of Michigan, 1996.
3. H. R. Everett y Douglas W. Gage. From Laboratory to Warehouse: Security Robots Meet the Real World. *The International Journal of Robotics Research*, 18(7):760–768, 1999.
4. H. R. Everett y D.W. Gage. A Third Generation Security Robot,. En *SPIE Mobile Robot and Automated Vehicle Control Systems*, 2903, páginas 118–126. SPIE, Boston, MA, 21 noviembre 1996.
5. R. Marín, P. J. Sanz, y A. P. del Pobil. The UJI Online Robot: An Education and Training Experience. *Autonomous Robots*, (15):283–297, 2003.
6. José Mira. AVISA: Un sistema semiautomático de diagnóstico de situaciones en tareas de vigilancia basado en técnicas de Atención VIsual Selectiva y dinámica con capacidad de Aprendizaje. TIN2004-07661, 2004.
7. José Mira y Antonio Fernández Caballero. D1-Estado actual del proyecto AVISA. informe técnico, UNED-UCLM, 2005.
8. José Mira y Antonio Fernández Caballero. D2-Estado actual del proyecto AVISA. informe técnico, UNED-UCLM, 2005.
9. Robin R. Murphy. *Introduction to AI Robotics*. MIT Press, 2000.
10. Sascha A. Stoeter, Paul E. Rybsky, Kristen N. Stubbs, Colin P. McMillen, Maria Gini, Dean F. Hougen, y Nikolaos Papanikolopoulus. A robot team for surveillance tasks: Design and Architecture. *Robotics and Autonomous Systems*, (40):173–183, 2002.

De simbólicos vs. subsimbólicos, a los robots etoinspirados

José María Cañas y Vicente Matellán

Grupo de Robótica
Universidad Rey Juan Carlos, 28933 Móstoles (España),
{jmplaza, vmo}@gsyc.escet.urjc.es,
http://gsyc.escet.urjc.es/robotica

Resumen En la Inteligencia Artificial, desde sus orígenes, han existido dos corrientes básicas, la simbólica y la subsimbólica. Estas dos aproximaciones han tenido gran influencia también en la robótica. En este artículo queremos presentar un enfoque menos conocido, el de la etología, y en concreto su aplicación a la generación de comportamiento autónomo en robots móviles. Para ello presentamos los fundamentos de la "Jerarquía Dinámica de Esquemas", una arquitectura para el control de robots móviles, basada en la composición de unidades simples denominadas "esquemas" siguiendo las teorías etológicas de Arbib. Igualmente se presentan experimentos preliminares que validan esta aproximación y se discute su viabilidad y se presentan los trabajos previstos para continuar investigando en esta línea.

1. Inteligencia artificial y robótica

En 1950 el Dr. Grey Walter publicó "An imitation of life"[25] en el que describía su trabajo con las tortugas-robot[1] Elmer y Elsie que había construido durante 1949 y que pueden considerarse como los primeros robots móviles. Sus trabajos son por tanto prácticamente contemporáneos de los de Warren S. McCulloch y Walter Pitts, que publicaron en 1943 "*A logical calculus of the ideas immanent in nervous activity*" [18] y de los de Norbert Wiener, que publicó su trabajo publicó "Cybernetics"[26] en 1948.

No sin discusión, podríamos considerar que todos estos trabajos supusieron el nacimiento de la robótica, la cibernética, y como consecuencia de la convocatoria de 1995[2] de J. McCarthy, M. L. Minsky, N. Rochester, y C.E. Shannon, de la Inteligencia Artificial (IA en adelante).

[1] La historia completa de estos ingenios puede consultarse en http://www.ias.uwe.ac.uk/Robots/gwonline/gwonline.html

[2] La convocatoria de lo que hoy conocemos como Conferencia de Dartmouth tuvo lugar en 1995 en la forma de convocatoria de un estudio durante dos meses por 10 personas del supuesto de que la inteligencia podría ser descrita de forma tan precisa, que una máquina pudiera simularla.

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 332-344, 2006.

Simplificando mucho la compleja historia de la IA desde entonces, podemos decir que ha tenido y tiene dos familias básicas, la IA subsimbólica, interesada en modelar la inteligencia a un nivel similar al de la neurona, de forma que cosas como el conocimiento y al planificación "emergerán"; y la IA simbólica, que modela elementos como el conocimiento y la planificación en estructuras de datos que tienen sentido para los programadores que las construyen. Otra forma de explicar la diferencia ente ambas aproximaciones, es la fuente de su inspiración, la biología en el caso de la IA subsimbólica y la psicología en el caso de la IA simbólica.

La psicología cognitiva, que trata de entender el funcionamiento de la inteligencia humana, ha ejercido gran influencia sobre la IA, que a su vez persigue poder reproducirla en una máquina. En este sentido, quizá apoyado en un razonamiento introspectivo, el paradigma dominante en la IA ha sido el de la descomposición funcional de la inteligencia en módulos especialistas (lenguaje, visión, razonamiento, etc.) y la lógica como motor de la deliberación, que se plasma como cierto proceso de búsqueda dentro de las alternativas.

También su ascendiente sobre la robótica móvil ha sido muy intenso. Por ejemplo, aportando la idea de que el comportamiento se puede generar como la ejecución de cierto plan calculado de antemano. Dicho plan es fruto de una deliberación sobre cierta representación del mundo, que tiene en cuenta los objetivos del robot. Por ejemplo, una ruta como secuencia de tramos intermedios que acaban llevando al robot al punto destino. La IA enfatiza la planificación y el modelado de la realidad como ingredientes fundamentales de la inteligencia en los robots.

El robot Shakey [20], un pionero dentro la robótica móvil, construido en el Stanford Research Institute (SRI) es el exponente más conocido de de la influencia de la *Inteligencia Artificial Simbólica*. Tenía una cámara, motores y sensores odométricos. Era capaz de localizar un bloque en su entorno y empujarlo lentamente. El procesamiento de las imágenes se realizaba en un ordenador fuera del robot. Dentro del mismo instituto, un digno sucesor fue el robot Flakey, construido en 1984 con numerosos avances tecnológicos sobre Shakey, y que ya incluía en su interior ordenadores para realizar a bordo cualquier cómputo necesario para su comportamiento.

Los orígenes epistemológicos de este enfoque están en la filosofía cartesiana que considera el alma como el ente que decide el comportamiento [19]. Hunde sus raíces en el cognitivismo y representa toda una teoría del funcionamiento de la inteligencia. Era natural que la Inteligencia Artificial clásica buscara refrendo probando su capacidad de generar comportamiento inteligente en cuerpos robóticos. Su aporte a la robótica ha sido fundamental en el desarrollo de esta última, introduciendo ideas como el manejo de símbolos, la planificación y la jerarquía para abordar la complejidad.

La aproximación subsimbólica, a pesar de haber sido empleada desde los principios de la robótica (las tortugas de Walter por ejemplo), virtualmente desapareció hasta finales de los 80. Hasta entonces los robots desarrollados, fundamentalmente basados en la IA simbólica, eran lentos y poco robustos, incapaces

de realizar con soltura tareas aparentemente sencillas como navegar por un pasillo y reaccionar a obstáculos imprevistos. Motivado por estas limitaciones nació el paradigma reactivo, inspirado en las asociaciones estímulo respuesta típicas del conductismo. En él se genera comportamiento asociando una respuesta de actuación a ciertos estímulos sensoriales, bien directamente [6], bien a través de transformaciones sensorimotoras [2].

En los últimos años esta tendencia hacia la biología se ha intensificado y en particular hacia la etología, buscando claves que permitan simplificar y hacer más robustos los comportamientos de los robots [16]. Varios conceptos nacidos en biología como la homeóstasis se han propuesto como mecanismos para la selección de acción [23]. También los movimientos de las abejas, capaces de volver siempre a su colmena y orientarse con el sol, de las moscas equilibrando el flujo óptico en ambos ojos, sirven de ejemplo para las técnicas de navegación en robots. La percepción gestáltica y el uso de invariantes visuales perceptivas como en los picados de los cormoranes para pescar [19] puede facilitar el desarrollo de comportamientos en robots, en apariencia muy sofisticados. También son destacables los trabajos que reproducen artificialmente el comportamiento de una rana [9] o una mantis religiosa [3].

En esta línea es en la que se plantea la arquitectura para el control de robots autónomos que describimos en este artículo. JDE (Jerarquía Dinámica de Esquemas) es una arquitectura para la generación de comportamiento autónomo en robots móviles inspirada en las ideas de la etología. Se basa en la utilización de *esquemas*, que se describen en la siguiente sección, como unidades básicas. La organización de estos esquemas se realiza mediante jerarquías, que se analizan en la tercera sección. En la cuarta sección se describen algunos experimentos que muestran la organización dinámica de los esquemas que es una de sus características más destacadas. Finalmente, la quinta sección resume algunas de las características principales así como los futuros trabajos que prevemos.

2. Esquemas

El encapsulamiento de funcionalidad en pequeñas unidades que luego pueden ser reutilizadas ha sido una constante desde los primeros intentos de comprender el comportamiento. A lo largo de estos años se han ideado diferentes unidades: módulos, habilidades, agentes, esquemas, comportamientos básicos, etc. cada una con sus matices y peculiaridades. En concreto, la aparición de los esquemas como parte explicativa del comportamiento surge en el campo de la fisiología y neurofisiología. Su instalación en el campo de la robótica ha recibido el apoyo de muchos investigadores, entre los que se podría destacar a Michael Arbib y Ronald Arkin[3].

Dentro de JDE definimos *esquema* como un flujo de ejecución independiente con un objetivo; un flujo que es modulable, iterativo y que puede ser activado o desactivado a voluntad. Distinguimos entre *esquemas perceptivos* y *esquemas*

[3] `http://www.cc.gatech.edu/aimosaic/faculty/arkin/`

motores o de actuación. Los esquemas perceptivos producen piezas de información que pueden ser leídas por otros esquemas. Estos datos pueden ser observaciones sensoriales o estímulos relevantes en el entorno actual, y son la entrada para los esquemas motores. Los esquemas de actuación acceden a esos datos y generan sus salidas, las cuales son comandos a los motores o las señales de activación para otros esquemas de nivel inferior (nuevamente perceptivos o motores) y sus parámetros de modulación, tal y como se ilustra en la figura 1.

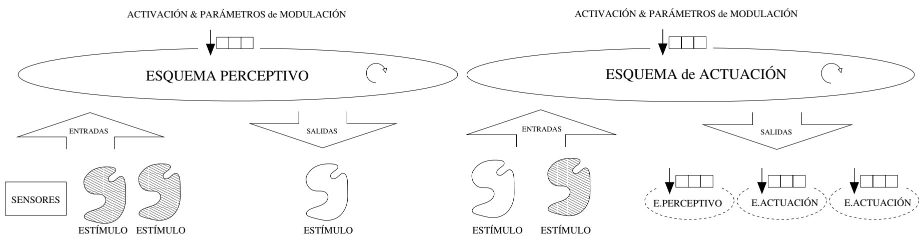


Figura 1. Patrón típico de un esquema perceptivo (a) y de un esquema motor (b) en JDE

Los esquemas JDE son por definición *modulables*. Siguiendo el segundo principio de Arbib [2], referente a la evolución y la modulación, los esquemas pueden aceptar varios parámetros de entrada que modulan su propio funcionamiento haciendo que se comporte de diferentes maneras. Además todos los esquemas en JDE son procesos *iterativos*, realizan su misión en iteraciones que se ejecutan periódicamente. De hecho, el periodo de esas iteraciones es un parámetro principal de modulación de cada esquema, permitiendo que se ejecute muy frecuentemente o con menor cadencia. Los controladores digitales se pueden ver como un ejemplo de este paradigma, pues ellos entregan una acción correctora cada ciclo de control. Los esquemas son además *suspendibles*, de manera que pueden ser desactivados al final de cada iteración, y en ese caso no producirán ninguna salida hasta que sean activados nuevamente.

A los esquemas se les asocia cierto *estado*[4], que ayuda a regular su coordinación, como veremos más adelante. Un esquema perceptivo puede estar en dos estados DORMIDO o ACTIVO. Cuando se encuentra ACTIVO, el esquema está actualizando las variables correspondientes al estímulo del cual se encarga. Cuando está DORMIDO las variables en sí existen, pero están sin actualizar, posiblemente desfasadas. El cambio de DORMIDO a ACTIVO o viceversa lo determinan los esquemas de nivel superior. Para los esquemas motores las cosas son un poco más elaboradas, pueden estar en cuatro estados: DORMIDO, ALERTA, PREPARADO y ACTIVO. Esto es debido a que cada esquema motor puede tener asociadas unas precondiciones y competir por el control con otros, según veremos en el apartado dedicado a la selección de acción.

[4] Estado en el sentido de los autómatas, es decir, que pasa de uno a otro

Con JDE el sistema completo está formado por una colección de esquemas, siguiendo el primer principio de Arbib sobre computación cooperativa de esquemas [2]: "Las funciones del comportamiento (perceptivo-motor) y la acción inteligente de animales y robots situados en el mundo se pueden expresar como una red de esquemas o instancias de esquemas que interactúan".

El empleo de este tipo de esquemas tiene dos implicaciones importantes: primero, la separación entre la parte perceptiva y la parte de actuación; y segundo, una fragmentación de ambas en pequeñas unidades que reciben el nombre de esquemas. La separación permite simplificar el diseño, porque la percepción y el control son dos problemas diferentes, relacionados pero distintos. Como ambos son complejos y distintos se resuelven en zonas del código separadas, eso facilita las cosas. La fragmentación en unidades pequeñas facilita la reutilización y permite acotar mejor la complejidad de cada uno de los subproblemas que aborda, haciéndolos manejables en una estrategia de divide y vencerás. Además esta separación permite dar cuerpo a cada esquema en un procesador distinto, posibilitando una implementación distribuida.

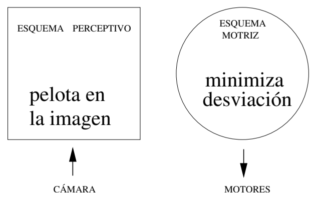


Figura 2. Comportamiento sigue pelota como activación simultánea de dos esquemas, uno perceptivo y otro motor

Esta división en esquemas motores y perceptivos se ha utilizado en [7,13,14,8]. Por ejemplo, en [13] se describen los comportamientos `sigue-pelota` y `sigue--pared` que se han conseguido como la conjunción de dos esquemas JDE. Tal y como muestra la figura 2 el comportamiento sigue pelota consta de un esquema perceptivo, mostrado con forma cuadrada, y un esquema motor, mostrado como un círculo (seguiremos esta notación a lo largo de todo el artículo: cuadrados para percepción y círculos para actuación). El perceptivo se encarga de buscar la pelota en la imagen y caracterizar su posición en ella, y el motor materializa un control proporcional que mueve el robot tratando de centrar la pelota en la imagen. Si la pelota aparece desviada a la izquierda, el esquema tratará de girar el robot hacia la izquierda, y de modo simétrico para desviaciones a la derecha. Igualmente, si la pelota aparece en la parte superior de la imagen, el esquema aumentará la velocidad de avance del robot porque la pelota está relativamente lejos. Si por el contrario aparece en la parte inferior significa que está demasiado cerca del robot, y el esquema hará retroceder al robot. Este esquema utiliza un control en velocidad sobre los motores de las ruedas y emplea la desviación de la posición central de la pelota respecto del centro de la imagen como error a minimizar. La ejecución simultánea de estos dos esquemas permite al robot generar la conducta observable de seguimiento de la pelota.

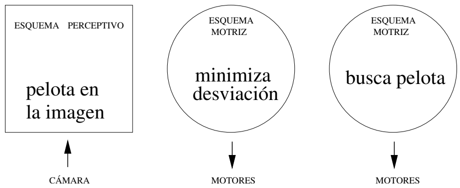


Figura 3. Sigue pelota con comportamiento apetitivo de búsqueda

Con estos dos esquemas, si no hay pelota alguna en la imagen entonces el esquema de control mantiene detenido los motores. Para ver la flexibilidad de esta descomposición, siguiendo con el sistema de ejemplo, una incorporación razonable podría ser un tercer esquema, también de control, que se activa cuando no aparece ninguna pelota en la imagen, y que se encargaría de mover el robot tratando de buscar las pelotas en su vecindad. Por ejemplo, girando el robot sobre sí mismo para barrer todos los alrededores. Tal y como muestra la figura 3, este esquema materializaría una conducta apetitiva, pues fomenta la aparición de la pelota en la imagen, que es el estímulo necesario para el comportamiento fundamental de seguimiento.

La pelota sería el estímulo clave que activa al esquema de seguimiento. En la figura 4 se añaden dos nuevos esquemas que enriquecen el comportamiento global con la capacidad de sortear obstáculos. Uno perceptivo para detectarlos y otro de control para sortearlos.

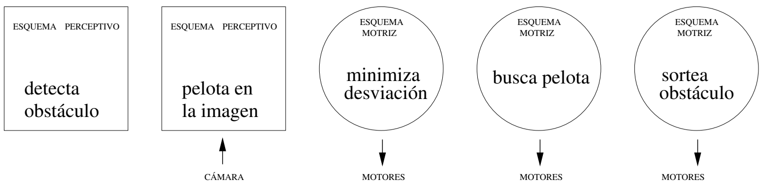


Figura 4. Sigue pelota complementado con sorteo de obstáculos

El siguiente paso sería englobar estos cinco esquemas como hijos de un esquema de nivel superior, que compite con otros, a su nivel, para hacer otra cosa, lo que introduce las jerarquías dinámicas que constituyen la principal aportación de JDE.

3. Jerarquía dinámica de esquemas

Una vez presentada nuestra unidad básica del comportamiento, el esquema, hay muchas opciones para su ensamblaje en un sistema completo. En este apartado daremos una visión global a la arquitectura JDE, que utiliza la jerarquía para regular el modo en el que se combinan los esquemas, y veremos también la manera en que éstos interactúan en ella y cómo unos utilizan la funcionalidad de otros para conseguir la conducta observable. Es decir, presentamos en esta

sección los mecanismos de percepción y actuación que ofrece esta jerarquía como principio organizador.

Tal y como vimos anteriormente, cada esquema de JDE tiene un objetivo propio, realiza alguna funcionalidad específica en la que es experto (bien sea en control, bien sea en percepción). La jerarquía aparece por el hecho de que los esquemas pueden aprovechar la funcionalidad de otros para materializar la suya propia, y en JDE el modo que un esquema tiene de aprovechar la funcionalidad de otro es precisamente su activación y modulación.

Este patrón de activación puede repetirse recursivamente, de modo que aparecen varios niveles de esquemas donde los de bajo nivel son despertados y modulados por los del nivel superior. Las activaciones en cadena van conformando una jerarquía de esquemas específica para generar ese comportamiento global en particular.

La colección de esquemas se organiza por tanto, en jerarquía para materializar cierta conducta. Esta jerarquía se reconstruye y modifica dinámicamente, según cambie el comportamiento a desarrollar o las condiciones por las que un esquema padre activa a cierto esquema hijo y no a otros.

La jerarquía que proponemos no es la clásica de activación directa, en la que en padre pone a ejecutar a un hijo durante cierto tiempo para que realice cierta misión, mientras él espera el resultado. En vez de que la misión del hijo sea un paso en el plan secuencial del padre, se entiende la jerarquía como una coactivación que expresa predisposición. En JDE un padre puede activar a varios hijos a la vez, porque no es una puesta en ejecución en la cual los hijos emitan directamente sus comandos a los actuadores, sino un situar en alerta. La activación final se deja en manos del entorno y de la competición con otros hermanos.

La forma en que se implementa ese proceso de activación y selección es usando cuatro estados: DORMIDO, ALERTA, PREPARADO y ACTIVO. El paso de DORMIDO a ALERTA lo determina la preactivación del padre. El paso de ALERTA a PREPARADO lo determinan las precondiciones del hijo, que él mismo evalúa periódicamente. El paso de PREPARADO a ACTIVO se pelea en una competición por el control entre los hermanos preparados en ese nivel.

En cada instante hay varios esquemas en ALERTA por cada nivel, ejecutándose concurrentemente, pero sólo uno de ellos es activado por la percepción del entorno. Cuando ningún esquema o más de uno quiere ser activado, entonces el esquema de nivel superior se invoca para que arbitre cual de ellos toma realmente el control. Por ejemplo, la figura 5 muestra una instantánea de la arquitectura JDE en funcionamiento. Los esquemas motores en ALERTA aparecen como círculos con bordes continuos. Sería el caso de los esquemas 5, 6 y 7. Varios de estos pueden ser compatibles con la situación perceptiva del entorno, pasando a PREPARADO, pero sólo uno de ellos ganará la competición por el control en ese nivel y pasará a ACTIVO. Los esquemas en ACTIVO se muestran en la figura 5 como círculos rellenos, como el 1, el 6 y el 15.

Esta idea de predisposición ya aparece en las propuestas de jerarquía aparecidas en etología. Además esta interpretación es compatible con que se pongan en alerta varios esquemas que realizan la misma función, y dependiendo de la

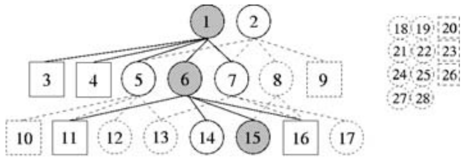


Figura 5. Jerarquía de esquemas y batería de esquemas en DORMIDO.

situación del entorno, sólo se activará el más adecuado a ese contexto. Tal y como señala Tinbergen para ejemplificar su jerarquía [22], el halcón que sale de caza tiene preactivados, predispuestos los módulos de `cazar-conejos`, `cazar--palomas`, etc. mientras sobrevuela su territorio de caza. Todos ellos satisfacen la finalidad de alimentarlo, se activará realmente uno u otro dependiendo de la presa que encuentre. Es lo que Timberlake [21] llama *variabilidad restringida*, y que tiene perfecta cabida en JDE. Esta capacidad de tener varias alternativas predispuestas de modo simultáneo contrasta con las alternativas que tiene RAP de Firby [10,11,12] (y por ello que hereda la arquitectura híbrida 3T [5]), que se ensayan una detrás de otra, pero sólo cuando la anterior ha fracasado.

Desde el punto de vista de los hijos, un esquema en estado activo sólo puede tener un único padre en un instante dado. Es decir, un hijo no puede tener varios padres en el mismo instante. Sí puede ocurrir que sea activado por distintos padres en diferentes momentos de tiempo. También puede suceder que se activen simultáneamente instancias diferentes del mismo esquema, probablemente con modulaciones distintas y en niveles diferentes. Esta posibilidad es un ejemplo de reutilización de esquemas.

Los esquemas que un momento dado no se usan para la tarea en curso descansan en una batería de esquemas, suspendidos en estado DORMIDO, pero preparados para la activación en cualquier momento. Éstos aparecen como cuadrados y círculos discontinuos en el lateral derecho de la figura 5 (esquemas 8, 9, 10, 12, 13, 17, 18, etc.).

Las jerarquías se construyen y modifican dinámicamente y son específicas de cada comportamiento. Si las condiciones del entorno o los objetivos del robot varían, puede ocurrir que entre los hermanos que están ALERTA en cierto nivel cambie la relación de cuáles de ellos están preparados para afrontar la nueva situación. Esto altera quién es el ganador en la competición por el control en ese nivel y fuerza una reconfiguración de la jerarquía desde ese nivel hacia abajo. Todos los que estaban activos por debajo se desactivan y el nuevo árbol se establece a partir del nuevo ganador. Por ejemplo, los esquemas 10, 12 y 13 de la figura 5 están a un sólo paso de la activación, pues serán despertados si el esquema 5 pasa a ACTIVO en su nivel.

Del mismo modo, si la situación del entorno o los objetivos se alteran, cierto esquema en determinado nivel de la jerarquía actual puede decidir modificar los parámetros de sus hijos actuales, o incluso cambiar de hijos. De nuevo, esto fuerza una reconfiguración de la jerarquía desde ese esquema hacia abajo, pues los

nuevos hijos al ejecutarse activarán a otros hijos suyos. Las reconfiguraciones en JDE suelen ser rápidas puesto que el arbitraje y las decisiones de cada esquema de control se toman periódicamente, a un ritmo suficientemente vivaz.

La jerarquía incluye la percepción como parte subsidiaria de los esquemas de actuación, y gracias a ello se facilita la coordinación, esto es, se da un contexto a la percepción. Para cada comportamiento hay que ligar los esquemas perceptivos a los de actuación, con ello se resuelve la coordinación percepción-actuación. Por ejemplo, la coordinación visuo-motora en el movimiento. La jerarquía determina qué percibir y cuándo.

Tal y como muestra la figura 5 hay reutilización de esquemas perceptivos puesto que el mismo esquema perceptivo puede asociarse a distintos esquemas motores. En realidad son distintas instancias del mismo esquema. Por ejemplo, el esquema 11 puede dar percepción al esquema motor 6 y al 5, o el 16 al 6, 7 y 8. Puede haber estímulos específicos, como el que elabora el 10, que sólo le interesa al esquema motor 5, pero en general la reutilización es la norma en percepción, para estímulos comunes. También puede haber reutilización de esquemas motores.

4. Experimentos con JDE

Validar una arquitectura para la generación de comportamiento autónomo es muy difícil. En nuestro caso hemos decidido que el único camino es la construcción de diversas aplicaciones, de las más simples a otras más complejas. Así hemos ido desarrollando diferentes aplicaciones como las descritas en [7,13,14].

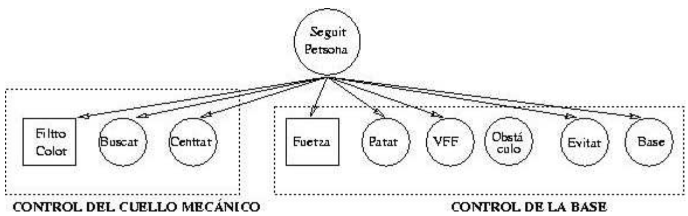


Figura 6. Jerarquía de esquemas para que un robot siga a una persona

Para ilustrar su uso vamos a describir brevemente dos aplicaciones. La primera es la reflejada en la figura 6. Se trata de la construcción del software para que un robot con visión monocular sea capaz de seguir a un humano en un entorno de interiores con desconocido, es decir, con todo tipo de obstáculos.

En esta aplicación, al ser sencilla, se puede resolver con un único nivel. En él se distinguen dos entornos de control diferentes (marcados mediante dos cuadrados punteados). Uno para el control del cuello mecánico (*pan-tilt*) y otro para el control de la base. El primero está formado por tres esquemas: buscar, centrar y filtrar. Los dos primeros motores y el tercero perceptivo. El segundo está formado

por un esquema perceptivo, encargado de proporcionar las fuerzas (denominado Fuerza en la figura) y cuatro motores para implementar la evitación de obstáculos, que detiene al robot y gira si aparece un obstáculo demasiado cerca, VFF para el control de robot con obstáculos a una distancia prudencial y el "Base" que se encarga de alinear la base con el cuello. Se ha añadido un esquema adicional de seguridad "Parar" para detener al robot al alcanzar el objetivo, o cuando se ha perdido. Se pueden encontrar los detalles de esta aplicación en [8].



Cuadro 1. Ejemplo de uso de JDE en un robot móvil

Las imágenes del cuadro 1 resumen el funcionamiento de la aplicación. Se puede observar el robot, una plataforma Pioneer fabricada por ActivMedia controlada por un ordenador portátil estándar y con una cámara de vídeo-conferencia. Las cuatro imágenes de la parte izquierda de la figura reflejan la evitación de un obstáculo, mientras sigue al móvil. Las cuatro de la derecha muestran el control separado del cuello mecánico y de la base, así se aprecia como inicialmente el seguimiento se realiza mediante el cuello mecánico, y como posteriormente la base se alinea con el cuello.

Un ejemplo de uso de JDE en el que se evidencia la necesidad de la jerarquía es el descrito en [17] en el que se construyó la jerarquía necesaria para controlar un robot simulado capaz de jugar al fútbol.

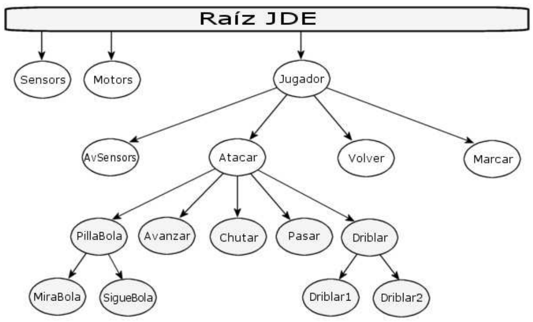


Figura 7. Jerarquía de esquemas para un jugador robótico de fútbol

La figura 7 refleja la forma en la que se organizan los esquema motores, en particular para un jugador ofensivo. Así, en el tercer nivel de la jerarquía, se puede observar como los esquemas "Chutar", "Pasar", "Avanzar", etc. estarán concurrente evaluando sus posibilidades, siendo las condiciones del entorno las que decidan cual de ellos será el que tome el control.

De esta forma, la construcción de comportamientos autónomos se simplifica enormemente, como en los sistemas deliberativos clásicos, pero sin perder además las ventajas de la reactividad de los sistemas basados en comportamientos. Todas estas consideraciones se analizan con mayor detalle en la siguiente sección.

5. Conclusiones y trabajo futuro

La idea central presentada en este artículo, y subyacente en JDE, es la creencia en la composición de comportamientos [1]. En nuestro caso entendiendo la composición en dos sentidos: temporal y en abstracción conceptual. En sentido temporal con JDE se argumenta que *una gran variedad de tareas motoras se puede describir y conseguir en términos de secuenciación de varios esquemas*, exactamente como enuncia el primer principio de Arbib [2]. En sentido conceptual, JDE apuesta por la ejecución simultánea de varios niveles de abstracción, cada uno de los cuales se materializa en otro nivel inferior y puede en cada momento materializarse en otro diferente de bajo nivel para adaptarse a las distintas condiciones del entorno.

Este planteamiento separa la abstracción conceptual de la abstracción temporal, que son indisociables en otras unidades de comportamiento como las unidades de acción [20] o los RAP [10]. En JDE los elementos de muy alto nivel de abstracción se pueden traducir en esquemas tan rápidos como se necesite. Se asciende en nivel de abstracción obligando a que las salidas de los esquemas de alto nivel sean exclusivamente activaciones de los esquemas de nivel inferior y su correspondiente modulación. Todos ellos funcionan en paralelo y toman decisiones en cada instante, cada uno en su nivel de abstracción. Un nivel más alto se distingue de otro inferior porque utiliza variables de estímulos que semánticamente son más abstractas, no por tener diferente velocidad.

La influencia de planteamientos etológicos en esta propuesta es muy grande. Por ejemplo, guarda muchas similitudes con la jerarquía que propone Tinbergen [22] para explicar la generación de comportamientos instintivos. Los esquemas de control son análogos a sus centros nerviosos. Como hemos señalado anteriormente, otra similitud es la utilización de la jerarquía para generar predisposición de ciertos comportamientos cuando se activa cierto instinto de nivel superior. Curiosamente las propuestas modernas de Arkin [4] también se apoyan en esta idea de jerarquía, donde se admite además la influencia que han ejercido en su sistema las ideas nacidas en la etología. Como valor añadido sobre el trabajo de Tinbergen y Lorenz, JDE explicita también la organización de la percepción, tal y como hemos visto.

JDE además permite la implementación directa de conceptos como los *estímulos clave* o los *comportamientos apetitivos*, cuya definición nació en el campo de la etología. Por ejemplo, el *comportamiento apetitivo* o de apetencia lo acuñó el biólogo Craig en 1926 y se utiliza para designar a aquella conducta que no satisface directamente ninguna necesidad del animal, sino que busca la situación desencadenante de la acción final, del *comportamiento consumatorio*, que es el que realmente sacia la necesidad interna [24,15]. Una muestra de conducta consumatoria podría ser comer una presa, mientras que buscarla sería apetitiva.

Los *estímulos clave*, por su parte, se definen como aquellos que disparan la activación de cierto patrón de comportamiento. En JDE se puede implementar de modo sencillo un esquema perceptivo que busca en la realidad el estímulo que desencadena el paso de ALERTA a PREPARADO de cierto esquema motor, promoviendo su activación. En general cada esquema motor tiene su *estímulo clave* [9], y reacciona ante él cuando está presente.

Estos conceptos "prestados"de la etología son los que se han conseguido en los experimentos descritos que validan la aproximación propuesta. El ejemplo de seguimiento permite certificar la capacidad de JDE para enfrentarse a entornos reactivos", siendo capaz de enfrentarse a todo tipo de cambios en el entorno. De igual forma, el ejemplo del jugador de fútbol permite validar JDE en lo referente a la orientación a objetivos, haciendo que el sistema, a lo largo del tiempo, vaya utilizando una serie de comportamientos básicos dependiendo de la situación, pero siendo finalista, es decir, cuando mejor contribuyen al objetivo general del jugador.

Finalmente, los trabajos futuros están orientados a aumentar la experimentación con jerarquía y escalabilidad. El experimento presentado en este artículo para justificar la jerarquía ha sido probado únicamente en un entorno simulado. Estimamos necesario evaluarlo sobre un robot real, trabajo que está actualmente en curso. Igualmente, se están aprovechando estos nuevos experimentos para probar una versión ampliada del código (denominada JDE+) implementada en C++ (la descrita en este artículo se realizó en C) y con nuevas funcionalidades. Igualmente, se están estudiando los mecanismos para emplear eficazmente la memoria a largo plazo y las técnicas de visión.

Agradecimientos

El trabajo descrito en este artículo ha sido parcialmente financiado por el proyecto ACRACE, financiado por el Ministerio de Educación y Ciencia (DPI2004-07993-03-01) y por el proyecto RoboCity 2030 (S-0505/DPI/000176) de la Comunidad Autónoma de Madrid.

Referencias

1. Eugenio Aguirre, María García-Alegre, and Antonio González. A fuzzy safe follow wall behavior fusing simpler fuzzy behaviors. In *Proceedings of the 3rd IFAC Symposium on Intelligent Autonomous Vehicles IAV'98*, pages 607–612, Madrid, March 1998.
2. Michael A. Arbib and Jim-Shih Liaw. Sensorimotor transformations in the worlds of frogs and robots. *Artificial Intelligence*, 72:53–79, 1995.
3. R. Arkin, A. Kahled, A. Weitzenfeld, and F. Cervantes-Prez. Behavioral models of the praying mantis as a basis for robotic behavior. *Journal of Autonomous Systems*, 32(1):39–60, 2000.
4. Ronald C. Arkin, Masahiro Fujita, Tsuyoshi Takagi, and Rika Hasegawa. An ethological and emotional basis human-robot interaction. *Robotics and Autonomous Systems*, 42:191–201, 2003.
5. R. Peter Bonasso, R. James Firby, Erann Gat, David Kortenkamp, David P. Miller, and Marc G. Slack. Experiences with an architecture for intelligent reactive agents. *Journal of Experimental and Theoretical AI*, 9(2):237–256, 1997.

6. Rodney A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1):14–23, March 1986.
7. José María Cañas, Marta Martínez, Pablo Bustos, and Pablo Bachiller. Overt visual attention inside JDE control architecture. In *Proceedings of the IROBOT Workshop inside Portuguese Conference on Artifical Intelligence EPIA2005*, 2005.
8. Roberto Calvo, José María Cañas, and Lia García. Person following behavior generated with JDE schema hierarchy. In *ICINCO, 2nd Int. Conf. on Informatics in Control, Automation and Robotics*, pages 463–466, Barcelona (Spain), sep 2005. INSTICC Press.
9. Fernando J. Corbacho and Michael A. Arbib. Learning to detour. *Adaptive Behavior*, 5(4):419–468, 1995.
10. R. James Firby. An investigation into reactive planning in complex domains. In *Proceedings of the 6th AAAI National Conference on Artificial Intelligence*, pages 202–206, Seattle, WA, 1987.
11. R. James Firby. Buiding symbolic primitives with continuous control routines. In *Proceedings of the 1st International Conference on AI Planning Systems AIPS'92*, pages 62–69, College Park, MD (USA), June 1992.
12. R. James Firby. Task networks for controlling continuous processes. In *Proceedings of the 2nd International Conference on AI Planning Systems AIPS'94*, pages 49–54, Chicago, IL (USA), June 1994.
13. Victor Gómez, José María Cañas, Félix San Martín, and Vicente Matellán. Vision based schemas for an autonomous robotic soccer player. In *Actas del IV Workshop de Agentes Físicos, WAF'2003*, pages 109–120, Universidad Alicante, March 2003. ISBN 84/607-7171-7.
14. David Lobato. Navegación local con ventana dinámica para un robot móvil. Proyecto fin de carrera, Universidad Rey Juan Carlos, February 2003.
15. Konrad Lorenz. *Fundamentos de la etología*. Ediciones Paidós, 1978.
16. Hanspeter A. Mallot and Matthias O. Franz. Biomimetic robot navigation. *Autonomous Systems*, 20:133–153, 1999.
17. Juanjo Martínez Gil. Equipo de futbol con JDE para la liga simulada robocup. Proyecto fin de carrera, Universidad Rey Juan Carlos, September 2003.
18. W.S McCulloch and W. Pitts. A logical calculus of the ideas immanent in neural nets. *Bulletin of Mathematical Biophysics*, 1943.
19. David McFarland and Thomas Bösser. *Intelligent behavior in animals and robots*. The MIT Press, 1993. ISBN-0-262-13293-1.
20. N.J. Nilsson. A mobile automaton: an application of artificial intelligence techniques. In *Proceedings of the 1st International Joint Conference on Artificial Intelligence IJCAI*, pages 509–520, Washington, (USA), 1969.
21. William Timberlake. Motivational models in behavior systems. In S.B. Klein R.R. Mowrer, editor, *Handbook of contemporary learning theories*, pages 155–209. Hillsdale, NJ: Erlbaum Associates, 2000.
22. N. Tinbergen. The hierarchical organization of nervous mechanisms underlying instinctive behavior. *Symposia of the Society for Experimental Biology*, 4:305–312, 1950.
23. Toby Tyrrell. An evaluation of Maes' "bottom-up mechanism for behavior selection". *Journal of Adaptive Behavior*, 2(4):307–348, 1994.
24. Güntel Vogel and Hartmut Angermann. *Atlas de biología*. Ediciones Omega, 1974.
25. Grey Walter. An imitation of life. *Scientific American*, 1950.
26. Norbert Wiener. *Cybernetics*. Wiley and Sons, 1948.

Arquitectura cognitiva para robots autónomos basada en la integración de mecanismos deliberativos y reactivos

J. A. Becerra, F. Bellas y R. J. Duro

Grupo de Sistemas Autónomos
Universidade da Coruña
ronin@udc.es, fran@udc.es, richard@udc.es

Resumen. En este artículo se propone una arquitectura cognitiva para robots autónomos que aúna las ventajas de las arquitecturas deliberativas en cuanto a la capacidad de aprendizaje en vida de modelos de entorno y de satisfacción, permitiendo su adecuada adaptación a procesos de aprendizaje guiado en entornos no estructurados, con las virtudes de las arquitecturas reactivas en términos de su inmediata reacción a los eventos que ocurren en entornos dinámicos. Esta integración se realiza por medio del establecimiento de un mecanismo de obtención de controladores reflejos siguiendo una arquitectura modular a partir de la interacción con el mundo de un mecanismo cognitivo darwinista.

1 Introducción

En el campo de la robótica autónoma existen diversas aproximaciones a la obtención de controladores para robots [1][2][3]. Estos controladores proporcionan el mecanismo para que el robot interactúe con su mundo y realice las tareas encomendadas. Ya sea de forma implícita o explícita, todas siguen un modelo utilitario basado en que para llevar a cabo cualquier tarea debe existir, una motivación que guíe el comportamiento en función del grado de satisfacción de la misma. Ésta puede explicitarse en el mecanismo y guiar el comportamiento o puede haberse asumido de forma implícita a la hora de obtenerlo.

En términos más formales, se asume que la percepción externa *e(t)* de un agente hace referencia a la información sensorial que es capaz de adquirir del entorno a través de sus sensores externos. Asimismo la percepción interna *i(t)* está constituida por la información sensorial que posee sobre su estado (propiocepción). Por lo tanto, la percepción global *G(t)* del agente estará formada por la percepción externa *e(t)* y la percepción interna *i(t)*.

Se establece también el concepto de satisfacción *s(t)* como el grado de cumplimiento de la motivación en un instante dado, que depende de la percepción global a través de una función de satisfacción S, y por tanto:

$$s(t) = S[G(t)] = S[e(t), i(t)]$$

Las percepciones externa e interna se pueden relacionar con la última acción realizada por el agente *A(t-1)*, con las percepciones externa e interna en el instante de tiempo anterior *e(t-1)* e *i(t-1),* y con los eventos externos e internos no controlados por el agente de tiempo característico inferior al modelable a través de las percepciones $X_e(t-1)$ y $X_i(t-1)$:

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 345-356, 2006.

$$e(t) = W [e(t-1), A(t-1), X_e(t-1)] \qquad i(t) = I [i(t-1), A(t-1), X_i(t-1)]$$

Si, como primera aproximación, despreciamos los eventos no controlados *X*, tenemos que la satisfacción del agente resulta:

$$s (t) = S [e(t), i(t)] = S \{[e(t-1), A(t-1)], [i(t-1), A(t-1)]\}$$

La interacción del agente con su entorno le debe llevar a la satisfacción de la motivación que, sin pérdida de generalidad, se puede expresar como la maximización de la función de satisfacción. Por tanto:

$$\text{máx} [s (t)] = \text{máx} \mathbf{(}S \{W [e(t-1), A(t-1)], I [i(t-1), A(t-1)]\}\mathbf{)}$$

La única variable susceptible de ser modificada por parte del agente en el proceso de maximización es la acción, ya que asumimos que la percepción externa y la percepción interna no se pueden manipular. Como consecuencia, un mecanismo cognitivo debe explorar el espacio de acciones posibles para maximizar la función de satisfacción. Esto implica la utilización de un algoritmo de búsqueda de extremos sobre la función *S* que, a su vez, depende de *W* e *I*. Previamente a poder llevar a cabo este proceso, debemos obtener las funciones *S*, *W* e *I* aplicando de nuevo un algoritmo de búsqueda de extremos pero, en este caso, sobre un espacio de funciones. Aquí se trata de maximizar la similitud de las funciones *I, W* y *S* con las funciones reales que representan el estado interno del agente, el entorno y la función de satisfacción del agente. Es decir, tratamos de modelizar la realidad del agente. Tendremos así cuatro algoritmos de búsqueda de extremos, tres en paralelo para encontrar las mejores funciones *W*, *I*, y *S* y un último que utiliza estas funciones para hallar la mejor acción. Las funciones *I* y *W* se suelen denominar modelo interno (*I*) y modelo de mundo (*W*).

Partiendo de esta estructura global, en la bibliografía existen distintas aproximaciones a su implementación dependiendo de cuáles de los elementos del esquema sean explícitos y cuáles se construyan en forma de "caja negra", así como de cuáles se precalculan antes de implantarlos en el robot y cuáles va obteniendo el robot de su interacción con el mundo. Por ejemplo, un sistema de control puramente reactivo establece una correspondencia directa entre percepciones en un instante dado e instrucciones a los actuadores, dejando implícitos los modelos de mundo, internos y satisfacción. Cuando se precalculan estas correspondencias directamente estamos ante una arquitectura reactiva estática. Esta aproximación es la que se sigue típicamente en los mecanismos basados en comportamientos más tradicionales [1] [3-5]. Evidentemente, si estos controladores en modo "caja negra" permitan adaptar su comportamiento en tiempo de ejecución (a través de algún mecanismo de refuerzo), tenemos una arquitectura reactiva adaptativa.

Por otra parte, uno podría generar explícitamente todos los módulos en tiempo real y calcular en cada instante la acción óptima en función de los modelos de que se dispone. Este otro caso extremo sería lo que se denominaría un mecanismo cognitivo deliberativo [2]. Existen, por supuesto, múltiples posibilidades intermedias que generan diferentes paradigmas [6-9]. A modo de esquema y, por relacionarlo con la terminología de los trabajos que encontramos en la bibliografía se puede presentar la clasificación de la Fig. 1.

Si se observa en la bibliografía, la mayor parte de los autores se circunscriben a una u otra aproximación, con algunos intentos de hibridar [10-12]. En este sentido se pueden plantear ventajas e inconvenientes de cada una, especialmente si hablamos de los extremos, esto es, las reactivas y los mecanismos cognitivos. En el caso de los mecanismos cognitivos tradicionales, se habla a menudo de aproximaciones basadas en conocimiento y en el procesado simbólico. Actualmente se pueden plantear aproximaciones, tales como el MDB (Multilevel Darwinist

Brain) [13][14] que permiten abordar y solucionar muchos de los problemas de las arquitecturas deliberativas tradicionales, especialmente en relación con el problema de *grounding* [15] al no trabajar con símbolos y al obtener sus representaciones directamente de la interacción del robot con el mundo a través de los sensores y actuadores. De todos modos, este tipo de arquitecturas no alcanzan los tiempos de reacción que permiten las arquitecturas puramente reactivas. Éstas a su vez imponen serias limitaciones a la capacidad de aprendizaje y adaptación en tiempo real en entornos complejos y variables que no son modelizables fácilmente. De hecho, la mayor parte de las arquitecturas reactivas se obtienen por diseño manual o por aprendizaje o evolución en entornos simulados [16-18], con el problema de la transferencia del comportamiento simulado a la realidad [19].



Fig. 1 Clasificación de los tipos de arquitecturas de control en robótica autónoma

El objetivo de nuestro trabajo es plantear una arquitectura que, de forma natural, pueda beneficiarse de las características positivas de las arquitecturas deliberativas darwinistas, en cuanto a su posibilidad de adquisición de conocimiento en tiempo real por interacción con el entorno, y la capacidad de reacción inmediata de las arquitecturas reactivas. Para ello, en este trabajo se plantea una arquitectura integrada basada en evolución que permite que las estrategias obtenidas por la arquitectura deliberativa MDB [13][14], cuando resultan exitosas puedan, de forma natural, irse convirtiendo en comportamientos reflejos en el agente que son activados en función del contexto. Para ello se opta por una redefinición del manejo de las acciones individuales que selecciona el MDB, que pasa ahora a seleccionar y almacenar estrategias en forma de controladores basados en comportamientos modulados [19].

2 Arquitectura deliberativa: MDB

El MDB (Multilevel Darwinist Brain) es un Mecanismo Cognitivo que permite a un agente autónomo decidir las acciones que debe aplicar en su entorno de cara a satisfacer sus motivaciones. El mecanismo se basa en una serie de teorías bio-psicológicas [20-22] que rela-

cionan el cerebro y su funcionamiento mediante un proceso Darwinista. A partir de estas ideas se ha construido un Mecanismo Cognitivo utilitario que utiliza los conceptos generales antes mencionados. Además, se han añadido nuevos elementos, como los pares acción-percepción, que son conjuntos de datos sensoriales y de actuación reales en un instante de tiempo, y que constituyen la información fiable y real de la que dispone el mecanismo para tratar de obtener los modelos.

En la Fig. 2 se muestra un diagrama funcional del MDB con los bloques básicos que lo componen unidos mediante flechas que indican el flujo de ejecución del mecanismo. El objetivo final es obtener la acción que debe ejecutar el agente en su entorno para satisfacer sus motivaciones. Esta acción se representa por medio del bloque marcado *Acción Actual* y se aplica al *Entorno* a través de los *Actuadores* obteniendo nuevos valores de *Sensorización* para el agente. Así, tras cada interacción con el mundo real, tenemos un Par Acción-Percepción que refleja la relación entradas-salidas, esto es, lo que percibió y actuó el agente y a qué percepción le ha llevado, que es lo que los modelos deben predecir.



Fig. 2. Diagrama de bloques del MDB

Estos pares Acción-Percepción se almacenan en una *Memoria a Corto Plazo* (MCP) y a continuación comienzan los procesos de *Búsqueda de los Modelos de Mundo, Modelos Internos y Modelos de Satisfacción* que mejor predigan los contenidos de la MCP. Estos tres procesos son concurrentes. Tras su finalización, los mejores modelos de cada tipo se marcan como *Modelo de Mundo Actual, Modelo Interno Actual y Modelo de Satisfacción Actual* y se utilizan en el proceso de *Optimización de la Acción* como entorno virtual en el que probar posibles acciones y determinar su efecto sobre la satisfacción. Así la acción se escoge en un proceso interno al MDB. Por tanto los modelos representan el conocimiento del entorno y de sí mismo que posee el agente a partir de la experiencia pasada.

La acción seleccionada se aplica en el *Entorno* a través de los *Actuadores* obteniendo nuevos valores de *Sensorización* y así nuevos pares Acción-Percepción que se almacenan en la MCP. Este ciclo básico se repite (lo denominamos una *iteración* del mecanismo) y, con el paso del tiempo, la información real disponible es mayor y, en consecuencia, los modelos son más precisos y las acciones escogidas a partir de ellos más adecuadas.

Los procesos de búsqueda de los modelos representan el modo en el cual el MDB adquiere el conocimiento a partir de los hechos y son procesos de aprendizaje, no de optimización (buscamos la mejor generalización posible a lo largo del tiempo, no queremos minimizar una función de error en un instante dado). Por este motivo, la técnica de búsqueda que utilicemos deberá permitir incorporar al dominio del problema nueva información que va obteniendo en tiempo de ejecución.

Basándonos en las teorías bio-psicológicas que fundamentan el MDB los modelos se plasman en Redes Neuronales Artificiales y los procesos de ajuste de estas redes son Algo-

ritmos Evolutivos. Estos algoritmos permiten un proceso gradual de aprendizaje si controlamos el número de generaciones de evolución en cada iteración. De este modo, los modelos obtenidos para un contenido dado de la Memoria a Corto Plazo (MCP) son fácilmente ajustables para otro contenido posterior, es decir, nunca estarán excesivamente optimizados para una situación particular (si el entorno es dinámico). Para lograr este aprendizaje gradual, debemos mantener las poblaciones de modelos entre iteraciones del MDB, de modo que únicamente al comienzo de la ejecución del mecanismo los modelos serán aleatorios. La función de calidad de los modelos se obtiene como una función de error de las salidas que proporcionan éstos cuando se prueban todos los pares acción-percepción que están almacenados en la MCP en una cierta iteración.

La obtención de los modelos en el MDB resulta computacionalmente costosa, por lo que una vez que se han obtenido modelos que resultan adecuados tras su aplicación, es importante que puedan ser almacenados de cara a facilitar futuros procesos de búsqueda o incluso a ser aplicados directamente en el futuro. No debemos olvidar que una de las características básicas en el diseño del MDB ha sido su aplicabilidad en entornos reales y, por tanto, dinámicos, de modo que la reutilización del conocimiento adquirido resulta básica. Por este motivo se ha incluido un nuevo tipo de memoria, la *Memoria a Largo Plazo* (MLP) cuya función primordial consiste en almacenar aquellos modelos (de cualquiera de los tres tipos) que resultan adecuados en su aplicación (modelos que predicen la MCP con niveles de error bajos, de manera estable y prolongada). La gestión de las memorias es un elemento fundamental en el buen funcionamiento del MDB. En este sentido, se ha desarrollado un sistema de gestión de memorias [23] que muestra el alto grado de interrelación que existe entre MCP y MLP en un mecanismo cognitivo.

Resumiendo, el MDB es un mecanismo cognitivo deliberativo que por medio de procesos evolutivos conforma en tiempo real los modelos de mundo, internos y de satisfacción y los va usando para tratar de obtener la acción óptima a realizar en un momento dado. El MDB y su funcionamiento están descritos exhaustivamente en una serie de publicaciones [13][14][23]. Aquí lo que se pretende es proponer una extensión del mismo de forma que este conjunto de modelos obtenidos en tiempo real no sólo se puedan utilizar para obtener una acción puntual, sino para conformar una arquitectura reactiva que permita al agente realizar acciones a largo plazo. Además, si se proporciona un sistema de memoria a largo plazo análogo al que se ha planteado para los modelos, estos controladores pueden ser parcial o totalmente reutilizados cuando se presentan situaciones similares llevando a la implementación de una memoria de tipo operacional. La reutilización de controladores reactivos para la generación de acciones será posible si éstos se generan de una manera ordenada y siguiendo algún tipo de arquitectura que permita la combinación de módulos y la reutilización de partes así como la adaptación a circunstancias ligeramente distintas. Es por ello por lo que se aborda en la siguiente sección la propuesta de una arquitectura de este tipo y la descripción de su obtención.

3 Arquitectura reactiva de los controladores

Como ya se ha indicado, el principal objetivo de esta arquitectura es conseguir que el proceso de obtención de controladores, basado en procesos de evaluación simulada en los modelos de que disponemos, sea lo más automático posible y sea también intrínsicamente incremental, facilitando esta obtención en casos de complejidad creciente mediante la maximización de la reutilización de controladores previos. Una aproximación de este tipo pasa por la modularidad y, en este caso, por seguir el planteamiento de la obtención de modelos en el MDB, hemos

optado por módulos que implementan comportamientos y que están basados en redes de neuronas artificiales (RNA). Los módulos, de los que hay dos tipos, de actuación y de decisión, reciben como entradas valores de sensores, bien sean virtuales o reales. Las salidas de los módulos serán, para los módulos de actuación, valores a introducir en los actuadores, reales o virtuales, y para los módulos de decisión valores que indican qué módulo(s) del nivel inferior debe(n) ejecutarse (Fig 3).

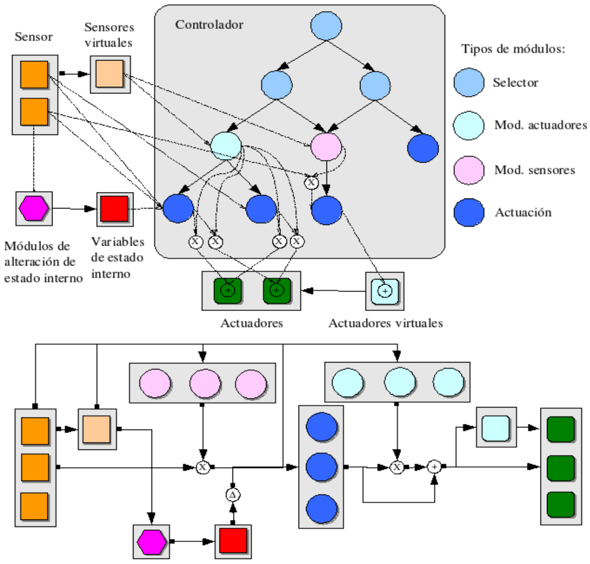


Fig. 3. Representación general para los controladores en la arquitectura

La idea inicial es que, para obtener un controlador que implemente un determinado comportamiento, se pueda partir de aquellos módulos previamente obtenidos, ya sean de actuación o de decisión, que se encuentren almacenados en una memoria. A partir de ellos se obtendrá un módulo que se situará en un nivel superior a los ya existentes y que decidirá en cada momento qué módulo o módulos de los ya disponibles tomará el control de los actuadores. Si, además, se necesita algún módulo de bajo nivel adicional, para realizar una actuación que ninguno de los módulos disponibles es capaz de proporcionar, se puede obtener simultáneamente con el módulo de alto nivel.

Respecto a la obtención, y por coherencia con el resto del funcionamiento del MDB, hemos optado por la utilización de algoritmos evolutivos. En definitiva, la obtención de un módulo consiste en un proceso evolutivo en el que los individuos representan controladores candidatos y son evaluados en un entorno para poder medir su calidad, es decir, su idoneidad. Ese entorno, en el caso que nos ocupa, está modelado a través de los modelos de mundo, interno y de satisfacción obtenidos por el MDB.

En cuanto al funcionamiento de los módulos de decisión, hemos establecido dos tipos distintos: selección y modulación. En la selección el módulo decisor escoge un único módulo de entre sus descendientes para su ejecución e inhabilita los restantes, mientras que en la modulación tiene la capacidad de activar varios de esos descendientes y cada uno en un grado distinto. En cuanto a la modulación, en este trabajo entendemos modular como “modificar levemente” la funcionalidad de un módulo de un controlador, esto es, la correspondencia que establece entre entradas y salidas. Formalizando un poco:

- Un módulo X es un antecesor de un módulo Y si hay un camino de X a Y.
- X es un descendiente de Y si hay un camino de Y a X.
- X es un descendiente directo de Y si hay un camino de longitud 1 de Y a X.
- X es un nodo raíz (R) si no tiene antecesores.
- X es un nodo de actuación (A) si sus salidas establecen valores para actuadores.
- X es un nodo selector (S) si su salida selecciona uno de sus descendientes directos para ejecución, cortocircuitando la ejecución de todos los otros descendientes directos.
- X es un nodo modulador de actuadores (AM) si sus salidas modifican (multiplicando con un valor entre 0 y 2) las salidas (no necesariamente todas) de sus nodos descendientes de tipo A. Las modulaciones se propagan a través del controlador hasta que alcanzan los nodos de actuación de tal forma que si entre un nodo R y un nodo A hay más de un AM que module una salida de ese nodo A, el valor final de modulación para esa salida será el producto de las modulaciones individuales presentes en el camino. Asumiendo que un nodo AM desea modular los valores de *n* actuadores, su número de salidas debe de ser necesariamente *n*numero de descendientes directos*, permitiendo así que la modulación que se propague por cada descendiente pueda ser distinta. Cuando existe más de un nodo A que proporciona valores para el mismo actuador, el actuador recibe la suma de esos valores.
- X es un nodo modulador de sensores (SM) si sus salidas modifican (multiplicando con un valor entre 0 y 2) las entradas (no necesariamente todas) de sus nodos descendientes. Las modulaciones se propagan a través del controlador hasta que alcanzan los nodos de actuación de tal forma que si entre un nodo R y un nodo Y hay más de un SM que module una entrada de ese nodo Y, el valor final de modulación para esa entrada será el producto de las modulaciones individuales presentes en el camino. Asumiendo que un nodo SM desea modular los valores de *n* sensores, su número de salidas debe de ser necesariamente *n*numero de descendientes directos*, permitiendo así que la modulación que se propague por cada descendiente pueda ser distinta.

En la parte superior de la Fig. 3 se muestra cómo sería un controlador con todos los tipos de módulos definidos. El controlador tiene comportamiento de árbol, aunque el árbol puede tener más de una raíz, pese a lo que parece en la figura donde existe un módulo con más de un antecesor. Eso se utiliza para indicar que hay módulos susceptibles de ser usados por más de un nodo, pero a la hora de ejecutar el controlador, si un nodo tiene *n* antecesores, se comportará como si estuviera replicado en las *n* ramas, pudiendo incluso ser ejecutado *n* veces con entradas distintas dependiendo del controlador en concreto.

Aunque el proceso constructivo nos proporcione un controlador con aspecto de árbol, éste se puede ver también como si únicamente tuviese dos niveles. Esto es así porque la selección no es más que una modulación que pone a 1 todos los actuadores de los nodos A del descendiente que activa y a 0 todos los actuadores de los nodos A de los restantes descendientes. La parte de debajo de la Fig. 3 muestra una representación general para los controladores en esta arquitectura teniendo esto en cuenta.

En definitiva, esta arquitectura permite la construcción progresiva por evolución de controladores complejos a partir de la reutilización de aquellos obtenidos previamente. Por lo tanto, implementando las acciones a realizar en el MDB como controladores de este tipo, convirtiendo el proceso de obtención de acción en la evolución requerida y añadiendo una memoria que permita conservar controladores que han resultado útiles, proporcionamos una vía para ir obteniendo progresivamente controladores reactivos en el seno del MDB que pueden llegar a ser utilizados incluso en forma de reflejos, esto es, sin una etapa deliberativa previa.

4 Ejemplos de operación

En este apartado pretendemos mostrar la idoneidad del MDB y de la arquitectura de modulación y su proceso de obtención para sus papeles en la arquitectura integrada propuesta. Inicialmente consideramos un ejemplo de la operación del MDB en un entorno real aprendiendo a realizar un comportamiento a través de las enseñanzas de un profesor. Con este ejemplo se mostrará las capacidades de aprendizaje y adaptación del mecanismo cognitivo MDB en sí. A continuación se presentará un segundo ejemplo donde se desarrolla una arquitectura de modulación reactiva para una tarea igual en términos de alcanzar un objetivo, por razones demostrativas hemos añadido una presa en el entorno para permitir observar los efectos de la modulación. Es la combinación de estas dos estrategias utilizando la arquitectura de modulación en la etapa de generación de acciones del MDB (en vez de optimizar acciones se optimizan controladores reactivos) lo que dota al sistema propuesto de las ventajas de ambas aproximaciones.

El robot Pioneer 2 utilizado en ambos ejemplos posee un anillo de sensores de sonar que permite calcular distancia a objetos cercanos y además le hemos incorporado un micrófono. Como actuadores posee dos ruedas y un brazo. En el primer experimento, el profesor guía al robot mediante un conjunto básico de 7 comandos que están codificados en forma de notas musicales y son captados mediante el micrófono. En función de la acción que selecciona el MDB, el robot se mueve. Dependiendo del grado de cumplimiento de la orden, el profesor premia o castiga al robot (mediante un valor numérico introducido directamente por teclado). Este premio también depende de la distancia recorrida por el robot en cada iteración y del ángulo final respecto al objeto. Este ciclo experimental básico se repite hasta que el robot alcanza el objeto. La motivación del comportamiento es simplemente recibir premios por parte del profesor, es decir, al robot le satisface que le premien. Una vez fijada la motivación, los modelos se construyen automáticamente en función de la sensorización disponible. La figura 4 muestra la representación concreta en este caso, donde se ha distinguido la operación con o sin profesor. Mientras el profesor está presente (modelos superiores de la figura), el modelo de mundo necesita únicamente dos entradas (el comando y la acción aplicada) y proporciona una predicción en la realimentación del profesor (es decir, si le premiará o no). El modelo de satisfacción es trivial en este caso, y tiene una única entrada (la predicción en la realimentación) a partir de la cual proporciona la satisfacción predicha.

El MDB funciona con todos los sensores del robot, no sólo con el micrófono y, por lo tanto, va obteniendo modelos de mundo con todos ellos. En el caso de los sensores de sonar, junto con la acción aplicada, constituyen el modelo de mundo que se muestra en la parte inferior de la Fig. 4. Sus entradas son la distancia al objeto, el ángulo de la parte frontal del robot al objeto y la acción aplicada. Las salidas están constituidas por los valores de sensorización predichos: distancia, ángulo. El modelo de satisfacción correspondiente proporciona de nuevo la satisfacción predicha a partir del incremento en distancia y del ángulo predicho por el modelo de mundo (parte inferior de la Fig. 4).

Mientras el profesor está presente, la acción que ejecuta el robot ha de ser seleccionada a través de los modelos disponibles, (los 2 superiores de la Fig .4). Si el profesor deja de dar órdenes (cuando considera que el robot ha aprendido de manera satisfactoria a obedecer), dejará de existir la entrada con el comando en el modelo de mundo de profesor, y, en este caso ese modelo no estará activo y su función será asumida por los otros modelos que llamaremos inducidos. El modelo de satisfacción inducido para los sonar ha ido aprendiendo que los incrementos en distancia y del ángulo obtenido en la operación con profesor tienen relación con la satisfacción, es decir, ha creado una relación entre la satisfacción del profesor y la posición

del robot respecto al objeto.

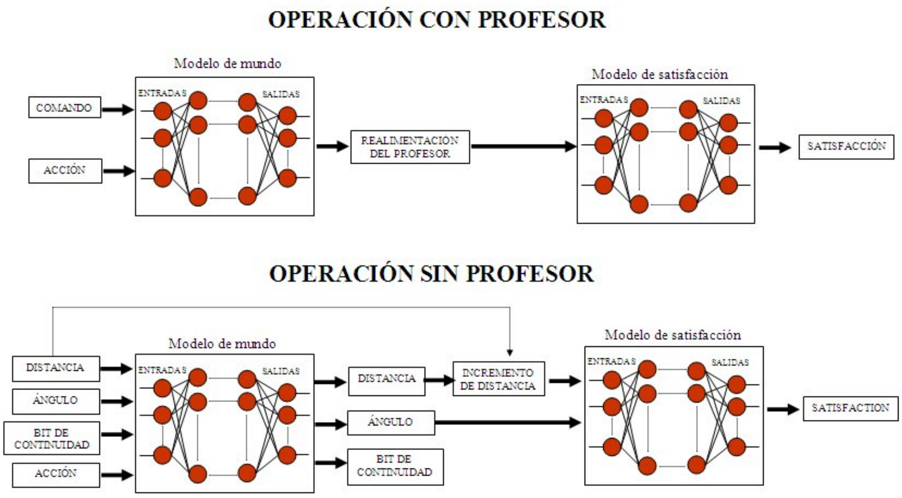


Fig. 4. Modelos de mundo y satisfacción para la operación del MDB con y sin profesor

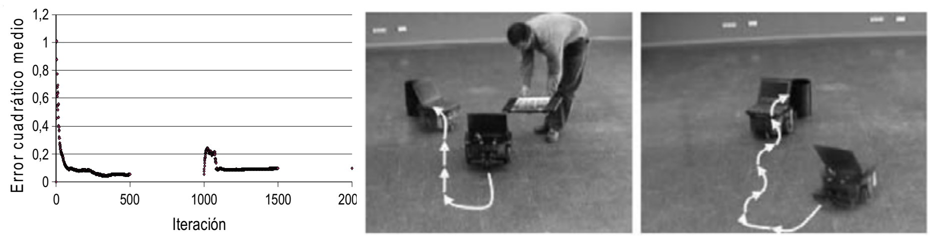


Fig. 5. Evolución del error cuadrático medio que proporciona el modelo de mundo actual respecto al contenido de la MCP en cada iteración del mecanismo (izquierda). La imagen central muestra el comportamiento del robot en la etapa con profesor y la imagen derecha el comportamiento inducido.

Los modelos utilizados en este ejemplo están representados mediante redes neuronales de tipo perceptrón multicapa simple y han sido evolucionadas con cuatro generaciones de evolución por interacción para lograr el aprendizaje gradual y general que comentamos en el apartado anterior. El tamaño de la MCP utilizada es de 40 pares acción-percepción. En la izquierda de la Fig. 5 mostramos la evolución del error cuadrático medio que proporciona el modelo de mundo actual respecto al contenido de la MCP en cada iteración. La etapa de operación con profesor abarca desde la iteración inicial (modelos aleatorios) hasta la iteración 500. Podemos observar en la figura como el error decrece rápidamente hasta un nivel inferior al 5%, que implica que el modelo es capaz de predecir de manera satisfactoria el contenido de la MCP. En la etapa inicial de aprendizaje de los modelos la tarea no es satisfactoria, pero rápidamente, según el error en los modelos baja, el robot comienza a capturar el objeto (escoge acciones de acuerdo con las órdenes del profesor) y esa tendencia no se pierde. En la iteración 500, el profesor deja de dar órdenes y las acciones se escogen mediante los modelos inducidos el número de capturas se mantiene constante. Este resultado implica que el MDB nos ha permiti-

do de una forma muy simple, obtener un comportamiento inducido pero sin que el diseñador haya predeterminado la forma de hacerlo. No existen datos en esta etapa sin profesor (desde la iteración 500 hasta la 1000) en la imagen izquierda de la Fig. 5 porque, al no existir comandos del profesor, el modelo de mundo asociado no evoluciona.

En este ejemplo se muestran las principales capacidades del MDB a nivel deliberativo, como son: capacidad de inferir modelos de entornos reales a partir de la interacción con el mismo y obtener acciones adecuadas a partir de ellos y la capacidad de modificar dichos modelos como respuesta al dinamismo del entorno. Hasta ahora las acciones estaban siendo seleccionadas individualmente por el MDB cuando una tarea de este tipo bien podría acabar siendo realizada por un controlador reactivo, sin el coste computacional derivado de la deliberación. Hemos comenzado a abordar este problema con objeto de lograr que el MDB sea realmente aplicable en robótica móvil. Deseamos que la etapa de optimización de la acción (Fig. 2) sea más compleja y permita que el MDB opere en un plano de abstracción superior, de modo que no tenga que enfrentarse con acciones de bajo nivel. Por este motivo, hemos aplicado la arquitectura reactiva e incremental en la etapa de selección de acción con el objetivo de que el MDB, en vez de seleccionar y evolucionar acciones, evolucione comportamientos en forma de controladores modulados, o sea acciones complejas a largo plazo basándose en la información de los modelos del MDB.

Como ejemplo de un controlador obtenido con la metodología comentada para la obtención de arquitecturas de modulación hemos planteado que el robot debe alcanzar a su presa evitando a un depredador. El esquema del controlador puede verse en la Fig. 6. Los dos módulos de comportamiento base son "escapar de depredador" y "cazar presa". El modulador de actuadores ajusta el funcionamiento de los módulos de actuación en función de lo cerca que esté de presa y depredador y en función del nivel de energía interno del robot, que a su vez depende de un hipotético consumo (toma mayores riesgos cuanta menos energía tiene ya que necesita con mayor urgencia "cazar y comer" a su presa). Finalmente, el modulador de sensores añade un nuevo factor a la ecuación, la ansiedad, de forma que ante baja ansiedad no altera el comportamiento, pero ante un nivel de ansiedad elevado el robot se arriesga en su estrategia de caza de la presa independientemente del nivel de energía interno. Cada módulo es un RNA que ha obtenido individualmente en procesos previos mediante evolución.

El comportamiento global ha sido obtenido en un entorno simulado formado por una habitación cuadrada, nuestro robot, el depredador y la presa. Se ha hecho asumiendo también velocidades constantes en los tres y comportamientos fácilmente predecibles del depredador (va en línea recta hacia nuestro robot) y de la presa (escapa en dirección opuesta). Este controlador es una perfecta base para que el MDB tenga un punto de partida sobre el que, en tiempo real, ajustar el comportamiento del robot al entorno en el que se está ejecutando en ese preciso momento, posibilitando así que aprenda a reaccionar correctamente ante estrategias más elaboradas por parte del depredador o la presa. En la Fig. 6 se puede apreciar el comportamiento en el robot real donde los papeles de depredador y presa los representan una papelera y una luz.

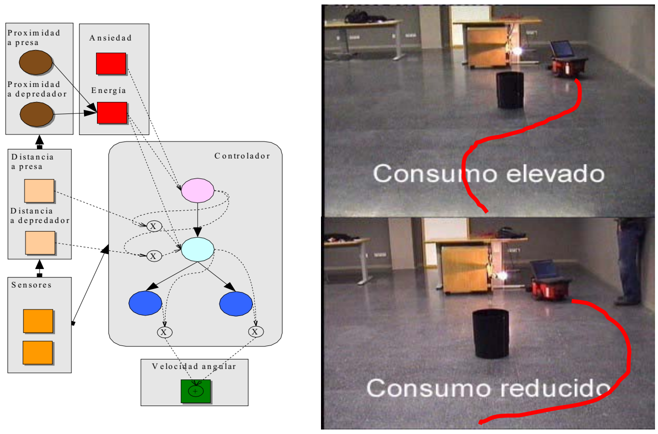


Fig. 6. Esquema del controlador obtenido con la arquitectura reactiva (izquierda) y comportamiento real obtenido

5 Conclusiones

En este trabajo se propone una arquitectura cognitiva para robots autónomos que aúna las ventajas de las arquitecturas deliberativas y reactivas. En este sentido se pretende obtener beneficio de la capacidad de aprendizaje en vida de modelos de entornos y de satisfacción, permitiendo su adecuada adaptación a procesos de aprendizaje guiado y entornos no estructurados que proporciona una arquitectura deliberativa darwinista, y de la velocidad de respuesta y bajo coste computacional de las propuestas reactivas. Para ello se propone un mecanismo de obtención de controladores reflejos siguiendo una arquitectura modular a partir de la interacción con el mundo de un mecanismo cognitivo darwinista (MDB). El MDB permite crear, en tiempo real y por interacción con el mundo, modelos de mundo, internos y de satisfacción que pueden ser utilizados como entornos virtuales en los que evaluar y evolucionar controladores o módulos de los mismos. Por otra parte, la arquitectura de modulación propuesta proporciona una estructura para permitir la creación de controladores complejos de forma incremental a través de un esquema de memoria y reutilización de soluciones que han sido adecuadas en contextos anteriores.

Agradecimientos

Este trabajo fue parcialmente financiado por el MEC a través del proyecto CIT-370300-2005-24 y la Xunta de Galicia a través del proyecto PGIDIT03TIC16601PR.

Referencias

[1] Maes, P. (1993), "Behavior–Based Artificial Intelligence", Proc. of the Second International Conference on Simulation of Adaptive Behavior (SAB92), pp. 2–10.

[2] Moravec, H.P. (1983), "The Stanford Cart and the CMU Rover", Proc. IEEE, V 71, 872–884.

[3] Brooks, R.A. (1986), A Robust Layered Control System for a Mobile Robot. IEEE Journal of Robotics and Automation, RA-2, pp. 14-23.

[4] Nolfi, S. (1997b), "Using Emergent Modularity to Develop Control Systems for Mobile Robots", Adaptive Behavior, Vol. 5, No. 3/4, pp. 343–363.

[5] Arkin, R.C. (1998), Behavior Based Robotics, MIT Press, Cambridge, MA.

[6] Nordin, P., Banzhaf, W., Brameier, M., (1998)"Evolution of a World Model for a Miniature Robot Using Genetic Programming", Robotics and Autonomous Systems Vol. 25, pp 105-116.

[7] Fukuda, T. and Hasegawa, Y. (2002), "Behavior Coordination and its Modification on a Monkey–type Mobile Robot", Biol. Insp. Robot Beh. Eng. Vol. 109, Phys–Verlag, pp. 45–87.

[8] R. Bonasso, R. Firby, E. Gat, D. Kortenkamp, D. Miller, M. Slack, "Experiences with an Architecture for Intelligent Reactive Agents", Journal of Experimental and Theoretical Artificial Intelligence, vol. 9, No. 2, 1997, pp. 237-256.

[9] Beer R., Quinn R., Chiel H., Ritzmann R., "Biologically Inspired Approaches to Robotics", Communications of the ACM, V.40 N. 3, pp 30-38, 1997.

[10] D. Lyons and A. Hendriks, "Planning as incremental adaptation of a reactive system", Robotics and Autonomous Systems, vol. 14, No. 4, 1995, pp. 255-288.

[11] E. Gat, "Integrating planning and reacting in a heterogeneous asynchronous architecture for mobile robots", SIGART Bulletin, Vol. 2, 1991, pp. 70-74.

[12] S. Koenig and M. Likhachev, "Improved Fast Replanning for Robot Navigation in Unknown Terrain", Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA), 2002, pp. 968-975.

[13] F. Bellas, A. Lamas, R. J. Duro (2001), "Adaptive Behavior through a Darwinist Machine", LNCI, vol 2159, pp 86-90, Springer

[14] F. Bellas, R. Duro, "Multilevel Darwinist Brain in robots: initial implementation", Proceedings of the ICINCO 2004 conference, 2004, pp 25-33., vol 2.

[15] Harnad, S. (1990), "The Symbol Grounding Problem", Physica D, Vol. 42, pp. 335–346.

[16] Ishiguro, A., Otsu, K., Fujii, A., and Uchikawa, Y. (2000), "Evolving an Adaptive Controller for a Legged-Robot with Dynamically-Rearranging Neural Networks", Proc. Supp. 6th Int. Conf. on Simulation of Adaptive Behavior, , pp. 235-244.

[17] Dorigo, M. and Colombetti, M. (1993), "Robot Shaping: Developing Autonomous Agents through Learning", Artificial Intelligence, Vol. 71, pp. 321–370.

[18] Kodjabachian, J. and Meyer, J–A (1995), "Evolution and Development of Control Architectures in Animats", Robotics and Autonomous Systems, Vol. 16, pp. 161–182.

[19] Becerra, J. A., Santos J., Duro R. J. (2003), "Multimodule ANN Architectures for Autonomous Robot Control Through Behavior Modulation", LNCS, vol. 2687, pp. 169-176.

[20] J. Changeux, P. Courrege, A. Danchin (1973) "A Theory of the Epigenesis of Neural Networks by Selective Stabilization of Synapses", Proc.Nat. Acad. Sci. USA 70, pp 2974-2978.

[21] M. Conrad (1974) "Evolutionary Learning Circuits". Theor. Biol. 46, pp 167-188

[22] G. Edelman (1987), "Neural Darwinism. Theory of Neuronal Group Selection". Basic Books.

[23] F. Bellas, J.A. Becerra and R.J. Duro (2006), "Construction of a Memory Management System in an On-line Learning Mechanism", Proc. ESANN 2006.

Modelización cualitativa para integración pluri-sensorial en un robot AIBO

David A. Graullera[1], Salvador Moreno[1] y M. Teresa Escrig[2]

[1] Instituto de Robótica, Universitat de València, Paterna, Valencia (Spain)
{david.graullera, salvador.moreno}@uv.es,
[2] Dpto. Ingeniería y Ciencia de los Computadores, Universitat Jaime I, Castellón (Spain)
escrigm@icc.uji.es

Resumen. Hasta la fecha, el problema de localización simultánea y construcción de mapas para navegación autónoma de robots móviles ha sido abordado principalmente mediante aproximaciones numéricas probabilísticas de alto coste computacional y bajo nivel de abstracción. Nuestro interés es utilizar una representación híbrida (cualitativa y cuantitativa) y modelos de razonamiento cualitativo para mejorar las limitaciones de la aproximación tradicional. En este artículo presentamos un método de integración pluri-sensorial de los sensores de un robot móvil (visión, infrarrojos y acelerómetros para el caso del robot AIBO) que permite capturar información del entorno de forma integrada y con un alto nivel de abstracción, tolerante a errores de sensorización, y que reduce las dependencias del conocimiento de control de los sensores particulares disponibles. Cada sensor tiene un módulo que incorpora su propio modelo cualitativo para actualizar el mapa híbrido del entorno, que integra de forma coherente el modelo del mundo más simple compatible con la información de los sensores proporcionada hasta el momento. La idea básica es que el modelo híbrido del mundo que posee el robot es utilizado, junto al conocimiento de las acciones emprendidas por el robot, para predecir las medidas o imágenes que los sensores deberían estar obteniendo, y el proceso de sensorización se limita a comparar estas predicciones con lo realmente observado. Mientras concuerden, el modelo del mundo se mantiene. Si hay una discrepancia, se dispara un procedimiento de unificación que genera hipótesis para explicar estas diferencias, según la cual se decide si la medida del sensor es incorrecta o si el modelo del mundo debe ser modificado, generando las propuestas de acciones exploratorias correspondientes, como visión activa, o repeticiones de medida.

1 Introducción

Las Bases de Conocimiento (BC) de los Sistemas Inteligentes clásicos para control en tiempo real de robots móviles autónomos se caracterizan por su estructura superficial, lo que las hace altamente dependientes de los sensores disponibles, y ocasionan problemas de generalización que obstaculizan su aplicación a diferentes tipos de robots móviles, en función de los sensores disponibles. Los sistemas expertos de segunda generación para robots móviles incluyen un modelo del ambiente en el que el robot se mueve, lo que les permite razonar sobre cómo obtener la información deseada del

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 357-364, 2006.

ambiente a partir del modelo del ambiente, y no directamente a partir de los sensores [Escrig, 05]. De esta manera, el conocimiento de control se vuelve independiente de los sensores, y es más robusto frente a la introducción de nuevos tipos de sensores, puesto que en vez de ignorarlos, o reescribir el conocimiento en función de los mismos, los nuevos sensores tan sólo mejoran la precisión del modelo del ambiente sobre el que se basa el conocimiento de control.

La aproximación propuesta consiste en una arquitectura para un sistema inteligente de segunda generación basada en la definición de una descripción híbrida cualitativa y cuantitativa del entorno en el que se mueve el robot, independiente de los sensores. Sobre este modelo, el sistema inteligente realiza dos tareas: una primera donde se integra la información sensorial a través de la modelización cualitativa de cada sensor con el objetivo de mantener el modelo híbrido del ambiente lo más parecido posible al estado real del ambiente en el que el robot se mueve, y una segunda donde el conocimiento de control realiza su objetivo basándose en el modelo híbrido del ambiente en vez de utilizar la información directa de los sensores.

2 Arquitectura

La arquitectura se compone de cuatro módulos: Los preprocesadores sensoriales cualitativos (QSPs), el modelo cualitativo (QM), la base de datos temporal (TDB) y el conocimiento de control (CK) (ver figura 1).

2.1. Base de datos temporal

La base de datos temporal (TDB) contiene la descripción cualitativa y cuantitativa del estado del ambiente en el que el robot se mueve en cualquier punto del tiempo, desde el pasado hasta el estado actual. Esta información, al ser de naturaleza cualitativa temporal, se representa mediante puntos temporales e intervalos entre los mismos según el cálculo de eventos de Kowalski y Sergot [Kowalski, 86]. El cálculo de eventos permite representar la evolución temporal de un sistema a partir de eventos que representan momentos del tiempo donde ciertos predicados sobre el sistema representado dejan de ser ciertos o comienzan a serlo, de forma que el tiempo queda dividido en una secuencia de eventos e intervalos entre eventos. En los intervalos entre eventos no puede ningún predicado hacerse verdadero o falso. Los eventos se convierten en las primitivas temporales como hechos prolog del tipo event(+Tiempo,+Objeto,+Lista_de_pares_predicado_valor), y las relaciones temporales son reglas que deducen cualquier pregunta a la TDB en base a los mismos. Si consideramos los estados cualitativos que describen el ambiente como los predicados representados, este modelo permite representar la evolución cualitativa temporal del ambiente de forma natural. Sin embargo existe la limitación de que, en un modelo híbrido, un estado cualitativo en un intervalo temporal puede implicar que una variable numérica del sistema cambie de valor de forma contínua entre dos eventos. Para ello podemos utilizar un predicado que afirme una ecuación de cambio progresivo, del estilo de (X is 30+4*T), que indica que la variable X cambia a lo largo del inter-

valo continuamente siguiendo una función del tiempo. Esto es lo que hace la modificación de cambio contínuo del modelo de Kowalski realizada por Shanahan [Shanahan, 90], y que hemos adoptado en nuestro modelo de TDB. Por último, la elección de variables cualitativas dependientes del espacio y del tiempo para describir el estado del ambiente fue realizada por los autores en el marco del proyecto EQUATOR, sobre razonamiento cualitativo y temporal para representación de sistemas de control de tráfico urbano [EQUATOR, 94]. En el trabajo actual, hemos extendido los predicados cualitativos para poder realizar razonamiento cualitativo espacial y describir así el ambiente, siguiendo el modelo de Freksa y Zimmermann [Freksa, 96] [Freksa, 00], según el cual se representan relaciones cualitativas de orientación y posición espacial entre puntos relevantes del ambiente, y se razona sobre el resto de puntos a partir de las relaciones conocidas.

Puesto que esta información es mucho mayor que la que realmente se ha obtenido de los sensores reales, debe completarse con hipótesis sobre la información faltante. La arquitectura propuesta se organiza como un conjunto de agentes inteligentes que interactúan a través de la TDB, por lo que ésta tiene una arquitectura de pizarra.

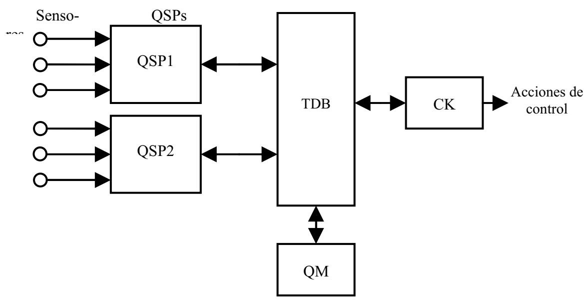


Fig. 1: Arquitectura propuesta

2.2. Conocimiento de Control

El conocimiento de control (CK) contiene el conocimiento no relacionado con la adquisición de datos, y constituye la parte de control del sistema experto. Utiliza la descripción cualitativa del estado actual del ambiente almacenado en la TDB como datos sensoriales. Puesto que esta descripción contiene información sobre los elementos que conforman el ambiente en cualquier punto del espacio y del tiempo (sea real o deducida como hipótesis), el conocimiento de control se mantiene independiente de los sensores disponibles.

2.3. Modelo Cualitativo

El modelo cualitativo predice el comportamiento del ambiente, teniendo en cuenta su estado actual y las acciones del robot que el conocimiento de control ha generado. El

modelo cualitativo genera un árbol de posibles comportamientos futuros del sistema, debido a la incertidumbre de la modelización cualitativa, de entre los cuales elegimos un futuro posible a través de la introducción de las suficientes asunciones como para fijar un único futuro. Estas asunciones están basadas en conocimiento heurístico y ordenadas por probabilidad de aparición. Pueden ser consideradas como la parte revisable del modelo, debido a que si en el futuro, se ve que este comportamiento futuro del modelo no es compatible con las entradas sensoriales, estas asunciones deben ser revisadas a través de un proceso de mantenimiento de la verdad. Debe además tenerse en cuenta que muchos modelos cualitativos carecen de una referencia temporal, por lo que se puede introducir una indeterminación innecesaria al comparar con los datos sensoriales recibidos en tiempo real, razón de mas por la que hemos optado por un modelo híbrido cualitativo-cuantitativo, tanto por la información cuantitativa temporal como por la espacial.

Generalmente, el estado actual instantáneo no puede ser obtenido directamente de los sensores debido a dos razones: primera, los sensores suelen estar integrados en periodos de tiempo, por lo que no existe información disponible durante ciertos intervalos de tiempo; segunda, los sensores sólo dan una pequeña parte de la información necesaria para comprender el ambiente en su estado actual. La aproximación propuesta es generar el estado actual del ambiente por simulación cualitativa a partir del estado anterior, en lugar de intentar analizar directamente los sensores. La simulación cualitativa desde el estado anterior dará una imagen completa del estado del ambiente asumiendo que el estado anterior es correcto, que el proceso de predicción es correcto, y que no han sucedido hechos impredecibles durante el intervalo de predicción. Los datos sensoriales son utilizados para verificar estos hechos, en lugar de generar el estado actual del ambiente por ellos mismos. El estado actual del ambiente generado por simulación cualitativa tiene mucha más información que la obtenida por los sensores, pero esto no es una paradoja: la mayor parte de esta información viene del estado anterior, y éste a su vez ha sido obtenido por integración de todos los datos sensoriales pasados y datos sobre la evolución del sistema.

2.4. Preprocesadores de sensores cualitativos

Los Preprocesadores de Sensores Cualitativos (QSPs) tienen el objetivo de mantener la evolución del modelo híbrido del ambiente lo más cerca posible de la evolución real del ambiente. Existe un QSP diferente para cada tipo de sensor. El QSP es un proceso dirigido por eventos, donde los eventos son producidos por la llegada de datos sensoriales. Cada vez que un dato sensorial es recibido, se inicia el QSP correspondiente, y su primer trabajo es consultar la TDB para comprobar si el dato es consistente con la evolución cualitativa almacenada en la TDB (siguiendo un esquema de representación del ambiente del robot cualitativo similar al descrito en [Falomir, 04]). Si es consistente, el QSP se detiene sin hacer nada. Si no lo es, el QSP revisa el conjunto de hipótesis tomadas por el QM para elegir el comportamiento futuro del ambiente almacenado en la TDB, entre el árbol de posibles futuros del sistema, para encontrar la hipótesis errónea. Se trata de un proceso de mantenimiento de la verdad. Finalmente se elige el mínimo conjunto de cambios de hipótesis necesarias para vol-

ver a mantener la consistencia con los datos recibidos, y se modifica la TDB en consecuencia, o se rechaza el dato sensorial por erróneo.

La figura 2 representa el proceso de actualización del QSP. El diagrama de la izquierda representa la evolución temporal real del sistema, y la derecha representa la descripción cualitativa del sistema en la TDB. Cada dato de sensor individual recibido por la QSP ha sido calculado de un área en el diagrama de la izquierda.

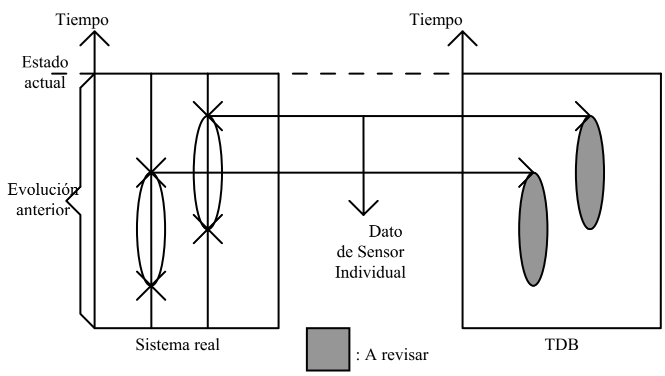


Fig. 2: Proceso de actualización del QSP.

Debe señalarse que la hipótesis de que el dato del sensor es correcto es también revisable. Si un sensor es defectuoso y ofrece datos erroneos, la revision de las hipótesis necesarias para que la TDB se vuelva de nuevo consistente con los datos sensoriales probablemente daría resultados tan improbables que la asumción más simple es descartar el dato sensorial y no tomarlo en cuenta en la TDB. Se consiguen así dos ventajas: primera, tolerancia a fallos de sensores o errores de reconocimiento sensoriales, puesto que el sistema los rechazará y su información será calculada sólo mediante simulación cualitativa, a partir del resto de sensores y el estado del modelo, por lo que la CK puede mantenerse independiente de los fallos sensoriales, y en segundo lugar, podemos indicar los sensores defectuosos en tiempo real, lo que puede ser interesante a efectos de mantenimiento.

3 Implementación

El sistema sobre el que hemos desarrollado esta arquitectura de integración plurisensorial es un robot cuadrúpedo AIBO de Sony[1]. Este robot tiene la forma de un perro, y sus sensores son principalmente una cámara de TV, dos micrófonos, tres sensores de distancia infrarrojos, varios acelerómetros, un sensor de vibración, otro de carga de la batería y unos sensores de tacto en el cuerpo, la cabeza y las patas. Nosotros nos hemos centrado para este trabajo en la cámara de TV, en los sensores infrarrojos y en los acelerómetros, dejando para un trabajo futuro la integración del resto de sensores.

El entorno de trabajo ha sido el siguiente: el robot se mueve en un laberinto con un

[1] www.eu.aibo.com

suelo de textura visual uniforme, aunque desconocida, y paredes planas, también de textura visual uniforme y desconocida a priori, aunque diferente al suelo y puede que diferentes entre sí. El área de movimiento del robot es por tanto una figura cerrada plana, compuesta por líneas rectas y esquinas.

Los programas se han implementado en ECLIPSE Prolog, que se conecta mediante sockets a un programa de interfaz realizado en C++ utilizando el AIBO Remote Framework RFW (http://openr.aibo.com), que a su vez se conecta al AIBO mediante un enlace inalámbrico configurado utilizando el AIBO WLAN manager. Las librerías visuales para Prolog se han implementado como programas C++ que se cargan en ECLIPSE como librerías dinámicas, para no sobrecargar el socket de Prolog con información visual.

El programa en Prolog tiene la forma de una arquitectura de pizarra con varios agentes interactuando con una pizarra. Estos agentes trabajan en paralelo concurrente, y la pizarra esta implementada como una base de conocimiento temporal (TKB), utilizando como representación el event calculus de Kowalski y Sergot modificado mediante el cambio continuo de Shanahan [EQUATOR, 94].

3.1. Base de Datos Temporal

La TDB contiene una descripción cualitativa de la evolución del ambiente que rodea al robot, de los sensores del robot y de sus actuadores. Todos los elementos en la TDB son de la forma:

```
event(+Tiempo,+Objeto,+Lista_Estado_cualitativo).

hevent(+Tiempo,+Objeto,+Lista_Estado_cualitativo).
```

Los event/3 representan eventos reales, que implican cambio en el estado cualitativo del objeto considerado, y entre dos eventos consecutivos del mismo objeto el estado cualitativo no cambia. Los hevent/3 representan hipótesis de evento, y son eventos supuestos por coherencia, pero no comprobados. Son eliminados o actualizados si se ve una contradicción con los datos sensoriales.

Para representar la información híbrida, se añade al estado cualitativo ecuaciones de cambio temporal según el modelo de cambio contínuo de Shanahan. Por ejemplo, un sensor infrarrojo que indica un acercamiento a una pared, se indicaría de forma similar a:

```
event(125,sensor_cabeza_ir_cerca,[cerca,(X is 48-4*T)]).

hevent(137,sensor_cabeza_ir_cerca,[choque,(X is 0)]).
```

Lo que indica que el sensor de infrarrojo detecta en t=125 que el estado a cambiado a cerca (menos de 50cm), y el estado híbrido indica que se prevee que evolucione según la ecuación X is 48-4*T, lo que permite elevar la hipótesis de que en t=137 chocará contra el obstáculo al que se esté acercando.

De la misma manera, la TDB contiene una descripción del estado cualitativo y temporal, junto con la información híbrida de cambio continuo, de cada sensor, de cada actuador del robot y de cada elemento conocido de su entorno. Esta representa-

ción es muy compacta, dado que únicamente se representan eventos significativos. Por ejemplo, en el ejemplo anterior, únicamente se indica el momento en el que el sensor pasa a indicar el estado cualitativo cerca, junto con una predicción de choque si no se hace nada por evitarlo, deducida de la distancia a la pared y la velocidad de movimiento del robot. Las sucesivas medidas del sensor no van a ser almacenadas a no ser que discrepen de las predicciones.

3.2. Preprocesador de sensor cualitativo para la visión cognitiva

El QSP para la cámara de TV utiliza procesado de video cualitativo para obtener una descripción cualitativa del laberinto desde el punto de vista de la cámara del AIBO. Las entradas de video son procesadas mediante una transformada de Gabor temporal [Bonet, 96], que es una transformación parecida a una wavelet cuyo núcleo consiste en ondas armónicas multiplicadas por gausianas, y que actúa en tres dimensiones, dos espaciales y una temporal, sobre la secuencia de video, por lo que puede detectar texturas en movimiento. Se trata de un modelo desarrollado en [Bonet, 96] a partir del modelo de córtex de Heeger. Las texturas móviles de superficie resultante son comparadas con las texturas predichas en la TDB a partir del conocimiento del movimiento del robot, y del mapa del entorno (compuesto por planos y esquinas, con indicación de las texturas de cada plano, y también de la textura del suelo), mediante un modelo híbrido de movimiento de superficies 3D con texturas uniformes en un plano de imagen. Si ambas concuerdan, la información de la cámara es ignorada, pues la TDB la puede reproducir. Si hay discrepancia, el QSP estudia varias hipótesis que pueden explicar la discrepancia, y elige la más probable.

3.4. Preprocesador de sensor cualitativo para los infrarrojos

El QSP para los telémetros infrarrojos debe predecir la distancia que éstos deben medir, y su velocidad de cambio, a partir del movimiento del robot y del modelo del mundo. Mientras coincida la medida de distancia generada con la predicha, no hacemos nada, pero si hay discrepancia, debe analizarse la causa. Generalmente, la causa de la discrepancia es un fallo momentáneo del sensor. Lo normal es descartar la información del sensor no concordante.

3.5. Preprocesador de sensor cualitativo para los acelerómetros

El QSP para los acelerómetros realiza una primera y segunda integración sobre los mismos. El modelo cualitativo predice, en función de los movimientos del robot, cuáles deben ser las medidas de los acelerómetros, y cuál debe ser la velocidad y la posición según las integraciones. Puesto que las medidas integradas ofrecen un error por la constante de integración, éste debe ser corregido reseteando la velocidad y aceleración del Aibo cuando este se encuentra inmóvil, según sus actuadores.

4 Conclusiones

La arquitectura propuesta ofrece varias ventajas:

- El diseño del Sistema Inteligente de Control (CK) se mantiene independiente de la evolución de los sensores, a la vez que se vuelve robusto frente a fallos sensoriales.
- La TDB contiene mucha más información sobre el estado actual del ambiente del robot móvil que los sensores. Esta información más rica simplifica el diseño de la CK.
- Los QSP pueden señalar e ignorar automáticamente datos de sensores cuya información no es consistente con la TDB, y que no pueden hacerse consistentes sin añadir a la TDB hipótesis muy difíciles de creer.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el proyecto CICYT con código TIC2003-07182.

Referencias

[Escrig, 05] Escrig, M.T., Peris, J.C, "The use of a reasoning process to solve the almost SLAM problem at the Robocup legged league", IOS-Press, Cat. Conference on Artificial Intelligence, CCIA'05, Oct. 2005.

[EQUATOR, 94] EQUATOR Final Report, ESPRIT II project, EEC 1994.

[Kuipers, 86] Kuipers, B.: "Qualitative Simulation", Artificial Intelligence, 29. pp. 289-338.

[Bonet, 96] E. Bonet, S. Moreno, F. Toledo and G. Martín: "A Qualitative Traffic Sensor based on Three-Dimensional Qualitative Modeling of Visual Textures of Traffic Behaviour", Proceeding of IEA-AIE 96. Fukuoka, Japan, 1996.

[Kowalski, 86] R. Kowalski M. Sergot: "A logic based calculus of events", New Generation Computing, pp. 67-96, Ohmaha ltd and Springer Verlag (1986)

[Shanahan, 90] M. Shanahan "Representing Continuous Change in the Event Calculus", 9th Proceeding of the ECCAI, Stockholm, Sweden (1990)

[Freksa, 96] C. Freksa, K. Zimmermann "Qualitative Spatial Reasoning Using Orientation, Distance, and Path Knowledge", in Applied Intelligence, Vol. 6., pag 49-58, 1996

[Freksa, 00] C. Freksa, R. Moratz and T. Barkowsky. "Schematic maps for robot navigation". In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Eds.), Spatial Cognition II - Integrating abstract theories, empirical studies, formal models, and practical applications (pp. 100-114), Berlin: Springer, 2000.

[Falomir, 04] Falomir Z., Escrig M. T., "*Qualitative multi-sensor data fusion"*, Seventh Catalonian Conference on Artificial Intelligence, IOS Press, Frontiers in Artificial Intelligence and Applications. Vol. 113, pp. 259-266, ISBN 1-58603-466-9, October 2004.

La arquitectura Acromovi: Una arquitectura para tareas cooperativas de robots móviles

Patricio Nebot y Enric Cervera

Departamento de Ingeniería y Ciencia de los Computadores
Universidad Jaume I, Castellón, España,
{pnebot, ecervera}@uji.es,
http://www.robot.uji.es/lab/plone/

Resumen Los avances en robótica móvil, poder de computación y comunicaciones inalámbricas han hecho posible el desarrollo de comunidades de robots autónomos. En los últimos años, hay un creciente interés en sistemas de múltiples robots autónomos capaces de llevar a cabo tareas cooperativas. Este proyecto tiene ésto en cuenta y presenta una arquitectura basada en agentes embebidos para el desarrollo de aplicaciones colaborativas para un equipo heterogéneo de robots móviles. La implementación de la arquitectura tiene la capacidad de permitir al equipo de robots el cumplimiento de tales tareas, y también tiene herramientas y características que permiten a los programadores desarrollar aplicaciones complejas en un tiempo razonable. Una plataforma de desarrollo basada en agentes ha sido usada para este propósito, debido a que integra las capacidades necesarias para el desarrollo de aplicaciones distribuidas, y maneja las comunicaciones entre todos los componentes.

Palabras clave: Sistemas multiagente, arquitecturas de control, coordinación de equipos de robots, cooperación, robots móviles

1. Introducción

Este artículo describe la implementación y desarrollo de una arquitectura distribuida para la programación y control de un equipo coordinado de robots móviles heterogéneos, los cuales son capaces de colaborar entre ellos y con personas para el cumplimiento de tareas de servicios en entornos cotidianos.

La presente arquitectura basada en agentes, Acromovi, nació a partir de la idea de que el trabajo en equipo es una capacidad fundamental para un grupo de múltiples robots móviles[1][2].

En los últimos años, hay un interés creciente en el desarrollo de sistemas de múltiples robots autónomos que puedan mostrar un comportamiento colectivo. Este interés es debido al hecho de que tener un único robot con múltiples capacidades puede ser una pérdida de recursos. Diferentes robots, cada uno con su propia configuración, forman un sistema más flexible, robusto y barato. Además,

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 365-376, 2006.

las tareas a realizar pueden ser demasiado complejas para un único robot, mientras que múltiples robots pueden realizar estas tareas de manera más efectiva[3].

La arquitectura Acromovi es un framework para el desarrollo de aplicaciones por medio de agentes embebidos y agentes interfaz con código nativo de bajo nivel. La arquitectura además implementa la compartición de los recursos de un robot entre todo el grupo. También facilita la cooperación entre los robots de manera que se puedan realizar tareas de modo coordinado. Además, Acromovi es una arquitectura distribuida que trabaja como middleware de otra arquitectura global para la programación de los robots.

Aunque la programación de robots se ha hecho mayoritariamente en C o C++, se ha elegido un sistema de desarrollo de sistemas multiagente basado en Java para el desarrollo de la arquitectura de nuestro equipo de robots[4]. Cualquier sistema de desarollo de sistemas multiagente común podría haber sido elegido, pero se ha optado por implementar la arquitectura por medio de JADE, una herramienta para el desarrollo de sistemas multiagente, implementada en Java, que cumple las especificaciones FIPA.

En la sección 2, se muestra el estado del arte relacionado con las arquitectura de control y sistemas de múltiples robots trabajando cooperativamente. En la sección 3, se describe la arquitectura Acromovi. Primero se verá el diseño lógico, y a continuación se mostraran los diseños e implementación de las dos capas que forman la arquitectura. Finalmente, la sección 5, muestra las conclusiones más relevantes obtenidas de nuestro trabajo.

2. Estado del arte

La investigación en el campo de la robótica móvil cooperativa ha crecido considerablemente en los últimos años[5][6]. Alguno ejemplos de estos trabajos son presentados a continuación.

En un intento de usar técnicas tradicionales de IA, la arquitectura *GOPHER* fue pensada para resolver problemas de una manera distribuida mediante multirobots en entornos internos[7]. Un sistema central de planificación de tareas (CTPS) comunica con todos los robots y dispone de una visón global de las tareas que han sido hechas o de la disponibilidad de los robots para realizar el resto de tareas.

Otro proyecto interesante, *MICRobES*, es un experimento de robótica colectiva que trata de estudiar la adaptación a largo plazo de una micro-sociedad de robots autónomos en un entorno con humanos. Los robots deben sobrevivir en este entorno así como cohabitar con las personas[8].

CEBOT (CEllular roBOTics System) es una arquitectura jerárquica descentralizada inspirada por la organización celular de las entidades biológicas. Es capaz de reconfigurarse dinámicamente para adaptarse a las variaciones del entorno[9]. Está compuesta por çeldas", y en la jerarquía hay çeldas maestras"que coordina las subtareas y se comunican entre ellas.

Con el fin de reutilizar componentes de software nativo, los agentes pueden ser embebidos en un nivel interfaz o middleware. *ThinkingCap-II*[10] es una

arquitectura desarrollada en un proyecto de plataforma distribuida basada en agentes para robots móviles. Incluye agentes híbridos inteligentes, sistema de planificación basado en tracking visual, componentes integrados de visión, y varias técnicas de navegación. Además, ha sido desarrollada sobre una maquina virtual en tiempo real (RT-Java), implementando un conjunto de comportamientos reactivos.

Por otro lado, *SWARM* es un sistema distribuido hecho de un gran número de robots autónomos. A partir de este proyecto, se ha acuñado el término "inteligencia SWARM"para definir la propiedad de los sistemas de múltiples robots no inteligentes que exhiben un comportamiento colectivo inteligente. Es una arquitectura homogénea en la cual la interacción está limitada a los vecinos más próximos[11].

Finalmente, *Miro* es un middleware para crear aplicaciones para robots móviles autónomos. Está basado en la construcción y uso de un middleware orientado a objetos para hacer el desarrollo de aplicaciones de robot móviles más fácil y rápida, y para fomentar la portabilidad y el mantenimiento del software del robot[12]. Este software además proporciona servicios abstractos genéricos, los cuales pueden ser aplicados en diferentes plataformas robóticas sin realizar modificaciones.

Los trabajos anteriores implementan arquitecturas y sistemas para que equipos de robots puedan realizar tareas cooperativas. Nuestro trabajo consiste en la implementación de una arquitectura de este tipo, resaltando la reusabilidad del software, permitiendo al programador la integración de componentes nativos de software (librerías de visión, módulos de navegación y localización). De este modo, los agentes embebidos son la clave para que el rápido desarrollo de potentes aplicaciones distribuidas.

También en este trabajo es importante resaltar otra característica de la arquitectura, La compartición de recursos de un robot entre todo el equipo, y el fácil acceso a los elementos físicos de los robots por parte de las aplicaciones.

3. Arquitectura ACROMOVI

Establecer mecanismos de cooperación entre robots implica considerar un problema de diseño del comportamiento cooperativo. Es decir, teniendo un grupo de robots, un entorno y una tarea a realizar, ¿cómo debe llevarse a cabo la cooperación? Este problema implica muchos retos, pero uno de los más importantes es la definición de la arquitectura que dará soporte a esta cooperación. Los sistemas multiagente son el entorno natural para tales grupos de robots, haciendo posible la rápida implementación de potentes arquitecturas para la especificación y ejecución de tareas.

Diseño de la arquitectura. Los robots que forman el equipo ya disponen de una arquitectura de programación con librerías nativas en C/C++ para todos sus accesorios. Esta arquitectura está formada por dos capas, la capa inferior, ARIA, se encarga de servir las peticiones de los programas a los distintos componentes

físicos de los robots. La capa superior, Saphira, proporciona servicios para la interpretación de sensores de rango, construcción de mapas, y navegación.

Sin embargo, el código nativo existente está orientado al procesamiento desde un único controlador, sin definir mecanismos de colaboración entre robots o aplicaciones. La arquitectura Acromovi trata de superar este problema mediante la introducción de una nueva capa sobre la arquitectura nativa que permita una fácil colaboración y cooperación.

Además, la arquitectura Acromovi es capaz de subsumir cualquier otro software extra, como ACTS o VisLib (librerías de visión). La subsunción de librerías C/C++ es un rápido y poderoso método para el desarrollo de código Java y por tanto agentes embebidos con toda la funcionalidad de las librerías nativas.

Como se puede ver en la figura 1, la arquitectura Acromovi añade una capa middleware entre la arquitectura del robot y las aplicaciones, que permite la colaboración y cooperación de los robots del equipo. Este middleware está implementado siguiendo un enfoque basado en agentes.

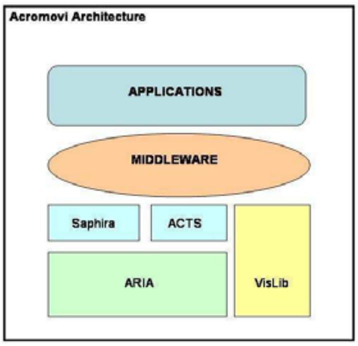
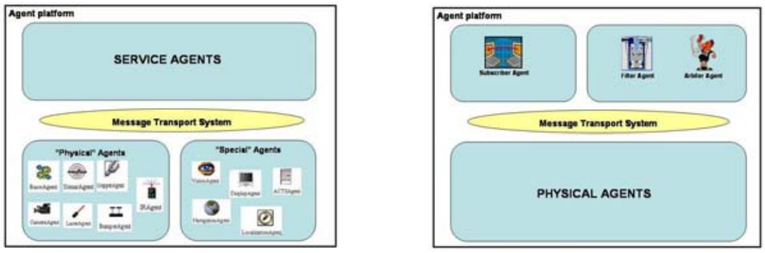


Figura 1. Diagrama general de la arquitectura

Como puede verse en la figura 2, la capa middleware ha sido dividida en dos subcapas. En la capa inferior, hay un conjunto de componentes que se encargan del manejo de los elementos físicos del robot, como el sonar, láser, base, ... El resto son componentes especiales que ofrecen servicios a la capa superior, como visión, navegación o localización.

Estos componentes solamente realizan dos tipos de tareas o funciones, el procesamiento de las peticiones que reciben y el envío de los resultados producidos. Así, éstos podrían haber sido implementados como componentes software, pero debido a razones de eficiencia, facilidad de implementación e integración con la estructura global, se decidió implementarlos como agentes que encapsulan la funcionalidad requerida.

La capa superior de la arquitectura comprende una gran variedad de agentes embebidos y de supervisión, como por ejemplo agentes para monitorizar el estado de los agentes de la capa inferior, agentes que proporcionan servicios de



(a) Capa inferior (b) Capa superior

Figura 2. La capa middleware

subscripción a un determinado componente, agentes que arbitran o controlan el acceso a un componente, y cualquier otro tipo de agente útil para la arquitectura o para una determinada aplicación. Además, estos agentes actúan como enlace entre las aplicaciones y el acceso a los componentes físicos del robot.

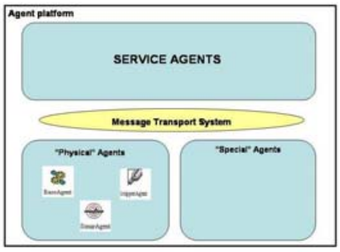
Cabe destacar que debido al carácter heterogéneo del equipo de robots y dependiendo de la configuración de cada robot en particular, pueden haber diferentes capas middleware para los robots. Estas capas middleware son generadas en tiempo de ejecución de acuerdo con los elementos que el robot tenga activos en el momento de su ejecución o puesta en funcionamiento. Estas diferentes configuraciones pueden verse en la figura 3.

Finalmente, sobre la capa middleware está la capa de aplicaciones, que también pueden ser implementadas como agentes. Estas aplicaciones, para acceder a los componentes físicos de los robots, deben comunicar con los agentes del middleware, los cuales luego acceden a la capa nativa que es la que controla a nivel físico el robot.

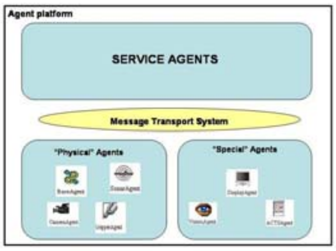
Una característica muy importante de la arquitectura creada es la escalabilidad de la arquitectura. Es decir, que la arquitectura puede crecer de una manera rápida y sencilla. Una vez una aplicación ha sido testada, si es útil para la arquitectura en general, ésta puede ser fácilmente convertida en un nuevo agente de la capa superior del middleware. A partir de ese momento, este nuevo agente se encargaría de ofrecer servicios, que puede ser de gran ayuda, a otras aplicaciones que se puedan crear. De este modo, cada aplicación que creemos puede incrementar nuestra arquitectura para poder crear aplicaciones cada vez más complejas y más interesantes, siguiendo un diseño "bottom-up".

Implementación de la arquitectura. Dado que el middleware es simplemente un conjunto de agentes embebidos, una herramienta para la programación de sistemas multiagente ha sido seleccionada para su implementación.

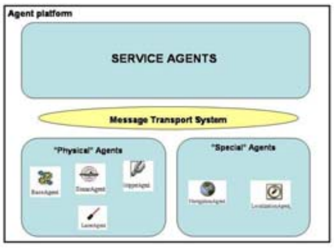
La herramienta elegida ha sido JADE (Java Agent DEvelopment Framework). Es un framework para el desarrollo de aplicaciones basadas en agentes de acuerdo con las especificaciones FIPA. Está totalmente implementado en Java y



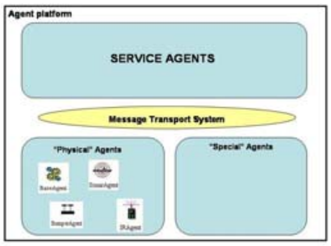
(a) Robot Base



(b) Robot con Cámara



(c) Robot con Láser



(d) Powerbot Base

Figura 3. Diferentes capas middleware.

simplifica la implementación de sistemas multiagente por medio de un middleware y un conjunto de herramientas gráficas que soportan las fases de desarrollo y depurado. El middleware de JADE implementa una plataforma para la ejecución de agentes y un framework para desarrollo de las aplicaciones. Además, proporciona facilidades para el agente como manejo del ciclo de vida, servicio de nombres y de páginas amarillas, transporte de mensajes y una librería de protocolos de interacción FIPA listos para ser usados[13].

La arquitectura Acromovi se ha implementado siguiendo las especificaciones de JADE. Así, cada robot involucrado en la arquitectura es un contenedor principal. Cada uno de estos contenedores funciona como host de una red distribuida. En cada uno de los contenedores hay un grupo de agentes. Estos agentes son creados en tiempo de ejecución dependiendo de la configuración del robot, como ya se dijo anteriormente. Cada uno de estos agentes representa uno de los elementos que el robot tiene activos en ese momento. Esto puedo verse en la figura 4.

Como se ha dicho, los elementos del robot están representados y manejados por agentes. Una breve explicación de los agentes que forman la arquitectura, tanto de la capa inferior como de la capa superior, se muestra a continuación.

Los agentes que forman la capa inferior del middleware han sido conceptualmente divididos en 3 grupos, dependiendo de la parte en la que el agente realiza su tarea. Así, hay "body agents"(base, gripper, sonar, bumper y IR agents),

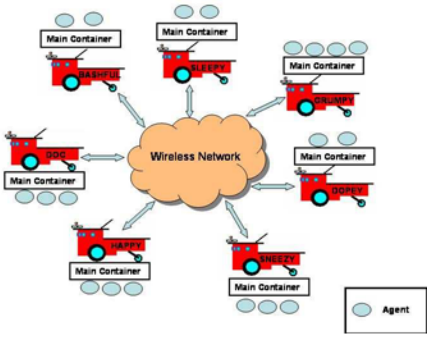


Figura 4. Estructura de la implementación

"laser agents"(laser, localization y navigation agents) y "vision agents"(camera, ACTS, vision y display agents).

Los "body agents"son los agentes necesarios o básicos para el funcionamiento del robot y corresponden con los principales componentes físicos de los robots del equipo. **Base Agent**, se encarga de todo lo relacionado con el movimiento de los motores y el estado interno del robot. **Gripper Agent**, agente encargado del buen funcionamiento de la pinza y de permitir a otros agentes la manipulación de ésta. **Sonar Agent**, encargado de que el sónar funcione correctamente y de devolver los valores apropiados cuando éstos son requeridos por otros agentes o aplicaciones. **Bumper Agent**, agente que maneja los bumpers del robot, y devuelve el valor de éstos cuando es requerido. **IR Agent**, agente dedicado al manejo de los 4 sensores de infrarrojos equipados en uno de los robots. Estos sensores tienen la función de prevenir al robot de escaleras o agujeros. El agente se dedica a avisar del estado de estos sensores cuando es necesario.

Los "laser agentsçomunican con el dispositivo láser y permiten el uso y manejo de dos módulos muy importantes para un robot móvil, la localización y la navegación. **Laser Agent** permite múltiples operaciones con el láser y se encarga de su correcto funcionamiento. **Localization Agent** se encarga de localizar al robot en un entorno conocido. **Navigation Agent** calcula el mejor camino a seguir entre dos puntos y mueve el robot por ese camino.

Los "vision agents"se encargan de capturar imágenes a través de la cámara del robot y realizar operaciones con esas imágenes. **Camera Agent** mueve físicamente la cámara e informa sobre su estado. **ACTS Agent** se usa para el tracking por color de objetos. **Vision Agent** captura las imágenes, y puede visualizarlas o modificarlas, mediante el uso de la librería Vislib. **Display Agent** modifica y procesa operaciones sobre las imágenes capturadas por el anterior agente, mediante el uso de las librerías Java 2D.

A continuación se describen los agentes de la capa superior. Estos agentes mantienen un carácter genérico, es decir, pueden tomar diferentes formas dependiendo del agente de la capa inferior al que sirven. Así, podría existir más de un agente del mismo tipo pero sirviendo a diferentes agentes de la capa inferior.

Estos agentes también han sido conceptualmente divididos en 2 grupos. Por un lado, hay un agente genérico encargado de controlar aquellos agentes de la capa inferior que pueden servir información o datos de una manera continua. El agente genérico encargado de ésto se llama **Subscriber agent**. Por otro lado, hay dos agentes encargados del control de los agentes de la capa inferior que puedan tener problemas de concurrencia en el acceso. Los agentes de la capa inferior candidatos a usar agentes de este tipo son aquellos que pueden mover físicamente alguna parte del robot, como puede ser la base, la pinza o la cámara. Los dos agentes encargados de su control son **Filter agent** y **Arbiter agent**. El funcinamiento específico de estos agentes se ve en profundidad en la siguiente sección.

4. Protocolos de interacción

Como muchos otros sistemas multiagente, Acromovi está compuesto por múltiples agentes interactivos. Estos agentes necesitan un modo de comunicarse entre ellos con el fín de alcanzar sus objetivos o aquellos de la sociedad en la que interactúan. Las herramientas que posibilitan esta comunicación entre los diferentes agentes de un sistema son los protocolos.

Un protocolo es simplemente un conjunto de reglas usadas para hacer posible una comunicación. En el dominio de los sistemas multiagente, un protocolo es un conjunto de reglas que guian la interacción que tiene lugar entre varios agentes y que gobiernan como la información es entregada. Estas reglas definen que mensajes son posibles para cada estado particular de interacción. Esa interacción tiene que ser realizada de tal forma que los agentes intercambien información. La interacción hace posible que los agentes puedan cooperar y coordinarse para conseguir realizar sus tareas.

Existen dos tipos diferentes de protocolos para ser definidos en un sistema multiagente, los protocolos de comunicación y los protocolos de interacción. Los protocolos de comunicación habilitan a los agentes para poder intercambiar y comprender los mensajes. Los protocolos de interacción se definen como series de actos comunicativos, formando conversaciones, para poder conseguir alguna forma de coordinación entre los agentes.

Como la arquitectura Acromovi ha sido implementada por medio de JADE, los protocolos de comunicación que se usan son los que proporciona, tales como TCP/IP, HTTP, RMI, ... La ventaja que ofrece JADE es que permite un uso transparente de estos protocolos. Además de que incorpora un mecanismo de transporte que se comporta como un camaleón, adaptando a cada situación el mejor protocolo de los disponibles, todo de forma transparente para el programador.

Por el otro lado, están los protocolos de interacción. Se utilizan los propuestos por FIPA, que además están implementados en las propias librerias de JADE. Algunos de estos protocolos son los protocolos Çontract Net", .English Auction.º "Dutch Auction". Pero estos protocolos resultan demasiado generales y no tienen en cuenta todas las características de la arquietctura Acromovi. Por tanto, se

han implementado dos nuevos protocolos, que añaden nuevas funcionalidades al sistema.

4.1. Protocolos de interacción

Los protocolos de interacción son series de actos de comunicación para conseguir alguna forma de coordinación entre los agentes. FIPA propone un conjunto de protocolos de interacción para sistemas de multiagente. JADE define una biblioteca con éllos, listos para usarse. Pero estos protocolos son generales y no tratan con algunas de las características del sistema. Por tanto, se han implementado dos nuevos protocolos que añaden nuevas funcionalidades al sistema.

El primer protocolo implica una modificación del protocolo de suscripción. En el protocolo de interacción FIPA para suscripción, un agente solicita tener notificaciones regulares de la información requerida. En el nuevo protocolo implementado, este enfoque cambia.

En este protocolo hay implicados dos tipos de agentes. Por una parte, los *suscriptores* son esos agentes que quieren tener un flujo regular de información de un elemento del robot, tales como los valores de las lecturas del sonar y el láser o un flujo continuo de imagenes provenientes del sistema de visión. Por otro lado, los *proveedores* son aquellos agentes que sirven los flujos de información. Estos agentes están estrictamente unidos a los agentes que manejan los elementos del robot.

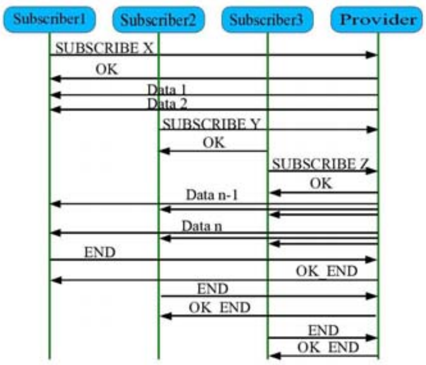


Figura 5. Protocolo de suscripción

Los agentes proveedores solicitan continuamente información a los agentes a los que están unidos. Así siempre tienen los valores actuales de estos agentes. La frecuenacia a la cual los agentes proveedores solicitan la información a los agentes controladores de los elementos viene dada por la frecuencia real a la que los elementos pueden servir la información.

El funcionamiento del protocolo puede verse en la figura 5. Los agentes suscriptores en el momento de la suscripción indican con qué frecuencia quieren

recibir los datos del agente proveedor (”SUBSCRIBE X”). El agente proveedor les sirve los datos a cada agente suscriptor en el momento adecuado (”DATA n”). Así, si un suscriptor quiere obtener los datos solamente cada dos veces que el sonar genera datos, el agente proveedor le manda al agente suscriptor los datos cada dos lecturas del sonar. El agente proveedor mantiene una lista activa en la que se indica el agente que ha hecho cada petición y a que frecuencia quiere obtener los datos.

El segundo protocolo implementado es completamente nuevo. Está pensado para controlar el acceso a los elementos problemáticos. ¿Cuáles son estos elementos problemáticos? Son todos esos elementos que pueden tener un funcionamiento erróneo si uno o más agentes pueden acceder y ejecutar operaciones al mismo tiempo en ellos. Estos elementos son la base, la pinza, la cámara.. Así, con este protocolo se intentan resolver los problemas de concurrencia. Otra característica importante de este protocolo es que previene que los agentes realicen operaciones con los elementos problemáticos que puedan poner en peligro al robot. Por ejemplo, si un agente quiere mover el robot dos metros, y hay una pared a un metro de distancia.

El funcionamiento conceptual del protocolo se puede ver en la figura 6. En este protocolo hay cuatro agentes implicados: el solicitante, el árbitro, el filtro, y el agente que maneja el elemento problemático. El agente solicitante es el agente que quiere utilizar el elemento problemático. Este agente debe solicitar permiso al agente árbitro para poder utilizar el elemento (REQUEST”). Así, cada agente que quiera utilizar un elemento problemático debe solicitar permiso al agente árbitro correspondiente.

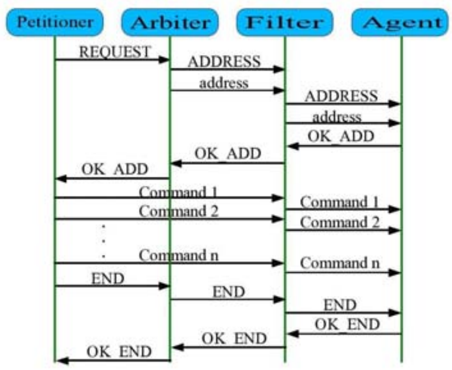


Figura 6. Protocolo de permiso

El agente árbitro mantiene una cola con las peticiones que han hecho los agentes solicitantes y va dando permiso siguiendo una estrategia FIFO. Cuándo el permiso es dado, el agente solicitante puede utilizar el elemento hasta que termine su tarea.

Cuándo el agente árbitro acepta la petición de un determinado agente, manda un mensaje con la dirección del agente solicitante al agente filtro indicando que agente que tiene el permiso (.ADDRESS”,.address”). El agente filtro, cuando recibe ese mensaje, manda un mensaje igual al agente que se encarga del funcionamiento del elemento, para informarle acerca de que agente tiene el control del elemento y a cuál debe mandar las respuestas de las órdenes ejecutadas.

Cuándo el agente que maneja el elemento recibe la dirección del agente solicitante, manda un mensaje al agente filtro confirmando que está listo para recibir las órdenes (.OK_ADD”). El agente filtro le manda otro mensaje al agente árbitro y finalmente, el agente árbitro le manda un mensaje al agente solicitante para confirmarle que tiene el permiso para utilizar el elemento.

Desde este momento, el agente solicitante puede mandar tantos mensajes como quiera hasta terminar su tarea (çommand n”). Los mensajes deben ser dirigidos al agente filtro, y éste los manda al agente del elemento si no son peligrosos para el robot. Así, el agente filtro se encarga de mantener la integridad de los robots. Si la orden mandada al agente filtro es conflictiva, el agente filtro no la manda al agente del elemento, y manda una notificación del error al agente solicitante.

Cuándo el agente solicitante termina su tarea, manda un mensaje al agente árbitro indicandole que ha terminado de usar el elemento (.END”). El agente árbitro manda un nuevo mensaje al agente filtro para informarle que el agente ha terminado y que debe esperar una nueva petición. El agente filtro manda un mensaje al agente del elemento para confirmarle que el agente solicitante ha terminado su trabajo. El agente del elemento manda un mensaje nuevo que confirma que ha recibido el mensaje al agente filtro (.END_OK”). Este le manda un mensaje al árbitro como confirmación. Y el árbitro le manda un mensaje al agente solicitante para confirmarle que ya no tiene el permiso.

Ahora, el agente árbitro toma la suiguiente petición de la cola o espera hasta que llegue una nueva.

5. Conclusiones

La principal conclusión que se puede obtener es que se ha logrado que la compartición de recursos de un robot entre el resto de robots del equipo se haga de una manera fácil y sencilla. Otra gran ventaja de la arquitectura es que permite la reutilización, mediante el uso de componentes nativos, por medio de agentes que proporcionan al programador un enorme conjunto de herramientas para robots móviles.

Por otro lado, se puede resaltar la escalabilidad del sistema, ya que una vez una aplicación ha sido testada, puede ser convertida en un nuevo agente e incrementar así la arquitectura.

Además, la arquitectura permite un rápido desarrollo de aplicaciones multirobot. En solo unas pocas semanas, estudiantes con bajos conocimiento de Java han sido capaces de desarrollar y programar agentes de diversas utilidades.

Por último, cabe destacar que diversas aplicaciones han sido implementadas para testar la arquitectura. Algunas de estas aplicaciones son el seguiemiento de robots por medio de visión, la navegación de robots usando la técnica del flujo óptico, y varias aplicaciones menores de depuración del sistema. Hay que remarcar que en todas estas aplicaciones, la arquitectura se ha comportado de la forma esperada, confirmando que es apta para la utilidad que se le planteó inicialmente.

Agradecimientos

Este trabajo ha sido financiado en parte a cargo del proyecto GV05/137 de la Generalitat Valenciana, y en parte a cargo del proyecto DPI2005-08203-C02-01 del Ministerio de Educación y Ciencia.

Referencias

1. Jung, D., Zelinsky, A.: An architecture for distributed cooperative planning in a behaviour-based multi-robot system. Journal of Robots and Autonomous Systems **26** (1999) 149–174
2. Mataric, M.J.: New directions: Robotics: Coordination and learning in multirobot systems. IEEE Intelligent Systems **13** (1998) 6–8
3. Arai, T., Pagello, E., Parker, L.E.: Guest editorial: Advances in multirobot systems. IEEE Transactions on Robotics and Automation **18** (2002) 655–661
4. Nebot, P., Gomez, D., Cervera, E.: Agents for cooperative heterogeneous mobile robotics: a case study. In: Proc. IEEE Int. Conf. on Systems, Man and Cybernetics. (2003)
5. Parker, L.: Distributed autonomous robotic systems 4. In: Proceedings of the 5th International Symposium on Distributed Autonomous Robotic Systems (DARS2000). (2000) 3–12
6. Cao, Y., Fukunaga, A., Kahng, A.: Cooperative mobile robotics: Antecedents and directions. Autonomous Robots **4** (1997) 1–23
7. Caloud, P., et al.: Indoor automation with many mobile robots. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems(IROS). (1990)
8. Picault, S., Drogoul, A.: The microbes project, an experimental approach towards open collective robotics. In: Proceedings of the 5th International Symposium on Distributed Autonomous Robotic Systems (DARS'2000). (2000)
9. Fukuda, T., Iritani, G.: Construction mechanism of group behavior with cooperation. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems(IROS). (1995)
10. Cáceres, D., Martínez, H., Zamora, M., Balibrea, L.: A real-time framework for robotics software. In: Int. Conf. on Computer Integrated Manufacturing (CIM-03). (2003)
11. Johnson, J., Sugisaka, M.: Complexity science for the design of swarm robot control systems. In: 26th Annual Conference of the IEEE Industrial Electronics Society (IECON). (2000)
12. Enderle, S., Utz, H., Sablatng, S., Simon, S., Kraetzschmar, G., Palm, G.: Miro: Middleware for autonomous mobile robots. In: IFAC Conference on Telematics Applications in Automation and Robotics. (2001)
13. Bellifemine, F., Caire, G., Poggi, A., Rimassa, G.: Jade a white paper. EXP in search of innovation (Special Issue on JADE) **3** (2003) 6–19

Desarrollo de un sistema inteligente de vigilancia multi-sensorial con agentes software

Juan Pavón[1], Jorge Gómez-Sanz[1], J. J. Valencia-Jiménez[2] y Antonio Fernández-Caballero[2]

[1] Universidad Complutense Madrid, Facultad de Informática
Ciudad Universitaria s/n, 28040 Madrid, España
{jjgomez, jpavon}@sip.ucm.es
[2] Departamento de Sistemas Informáticos, Escuela Politécnica Superior de Albacete e Instituto de Investigación en Informática de Albacete (I3A)
Universidad de Castilla-La Mancha, 02071 Albacete, España
caballer@info-ab.uclm.es

Resumen. Un sistema inteligente de vigilancia multi-sensorial consta de un conjunto variado de sensores y elementos fijos y móviles. Estos elementos son fuente de ingentes cantidades de información que tiene que ser contrastada y correlacionada para determinar si se producen situaciones especiales y actuar consecuentemente. Se trata asimismo de sistemas en entornos altamente dinámicos y con requisitos de seguridad y robustez muy exigentes. Por todo ello, es necesaria una solución distribuida, donde los elementos dispersos puedan decidir y actuar con autonomía (por ejemplo si quedaran aislados), cooperar y coordinarse para realizar un seguimiento completo de situaciones especiales. Para abordar esta problemática se ha optado por desarrollar todo el control del sistema de vigilancia como un sistema multi-agente.

1 Introducción

El desarrollo de nuevos tipos de redes inalámbricas y variedad de dispositivos sensoriales, captores y actuadores, con mayor capacidad de computación, permite la realización de sistemas de vigilancia cada vez más sofisticados [1, 2]. Estos sistemas constan de redes (cableadas o inalámbricas) de sensores (cámaras de video, micrófonos, detectores, etc.) [1] capaces de trabajar de forma omnidireccional y direccional (orientable en las tres dimensiones) [3], montados a su vez sobre unas plataformas móviles (artefactos motorizados que permiten el desplazamiento por las instalaciones) o fijas (ancladas en un determinado punto de la instalación) [4]. Una parte fundamental de este tipo de sistemas es su control.

Tradicionalmente, el control de un sistema de vigilancia se ha venido realizando de forma centralizada, con configuraciones donde los sensores reportaban a un controlador central, el cual tomaba las decisiones de qué hacer y transmitía instrucciones a actuadores remotos. Esta solución, cuyo diseño es conceptualmente sencillo (en cuanto a la arquitectura de sistema), tiene, sin embargo, varios problemas en cuanto a escalabilidad y robustez, que provienen de la rigidez jerárquica de la propia arquitec-

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 377-388, 2006.

tura. Por ejemplo, ante fallos o intrusiones en la red de comunicaciones puede haber áreas del sistema vigilado que quedarían descubiertas. O cuando se produce algún suceso grave pueden darse avalanchas de alarmas que colapsen el sistema de control dificultando su capacidad de decisión y reacción. Por todo ello es necesario considerar nuevas arquitecturas, más descentralizadas y más distribuidas. Esta distribución tiene dos aspectos principales. Por una parte, permitir que los distintos elementos del sistema tengan cierto grado de autonomía para poder tomar decisiones localmente y actuar independientemente del control central. De esta manera se pueden solventar problemas derivados del aislamiento temporal de dichos elementos y reducir las comunicaciones necesarias en el sistema, mejorando también la eficiencia global del sistema. Y por otra parte, hay que considerar la coordinación de los componentes del sistema distribuido. Esta coordinación permitirá mejorar el funcionamiento del sistema, por ejemplo, en la evaluación de la relevancia de los eventos que capturen los sensores, en el seguimiento de elementos móviles por el sistema vigilado, o para la colaboración de varios actuadores en la resolución de algún problema.

Teniendo en cuenta estas características, es razonable considerar la realización del sistema inteligente de vigilancia multi-sensorial como un sistema multi-agente. Los agentes son componentes software distribuidos, con autonomía para tomar sus propias decisiones y capacidad de percibir y actuar sobre su entorno. Dependiendo del grado de complejidad que requiera su comportamiento pueden tener una naturaleza puramente reactiva (esto es, la respuesta a los eventos que percibe se puede definir como una función más o menos sencilla) o cognoscitiva (pueden razonar e incluso aprender para adaptarse y proponer nuevas soluciones a un entorno cambiante). Asimismo, otro aspecto fundamental de los agentes es su sociabilidad, su capacidad para interactuar entre sí con el propósito de buscar soluciones coordinadas a los problemas que se plantean. Es por ello que se suele hablar de sistemas multi-agente (SMA).

La distribución de la inteligencia como en un SMA permitirá abordar las cuestiones que plantea el desarrollo del sistema inteligente de vigilancia multi-sensorial:

- Ancho de banda: El procesamiento distribuido evita la necesidad de un gran ancho de banda para transmitir la enorme cantidad de datos que producen los sensores hacia los nodos de procesamiento, teniendo en cuenta además que esos flujos de datos suelen ser altamente redundantes.
- Productividad: El procesamiento total del sistema aumenta al participar en él más nodos de forma paralela, por lo tanto se realizan más cómputos en el mismo tiempo que en la arquitectura centralizada.
- Velocidad: El procesamiento distribuido en los nodos sensoriales no solo incrementa el paralelismo del procesamiento total sino que también libera a los nodos de procesamiento centrales para concentrarse en cómputos y análisis más específicos que requieran de más recursos, eliminando trabas para que esos nodos centrales lleven a cabo sus tareas en menos tiempo.
- Robustez: La tolerancia a fallos se ve mejorada al poder replicarse componentes de control con lo cual se obtiene mayor redundancia, facilitando la reconfiguración y recuperación ante fallos. Asimismo, por la autonomía de los agentes, es posible ejecutar soluciones localmente ante determinado tipo de situaciones de fallo, sin depender de un controlador central. También, la coordinación entre los agentes permite mejorar los diagnósticos ante eventos que

capte el sistema.

- Autonomía: El sistema puede cubrir potencialmente una extensa área geográfica, por lo tanto es necesario eliminar el tráfico innecesario que pueden ocupar los canales de transmisión incrementando las probabilidades de congestión de tráfico, imposibilitando al resto de nodos periféricos comunicarse con los nodos centrales, lo que degeneraría en la caída del rendimiento global del sistema.
- Escalabilidad: El sistema puede crecer de forma más fácil y fiable, ya que la mayor parte de la carga de procesamiento es local a los nodos que han sido agregados, tanto si son nodos centrales de procesamiento como nodos periféricos.

Ya algunos sistemas de vigilancia han encontrado en la tecnología de agentes software y sistemas multi-agente el marco ideal para acometer la sofisticación de sus trabajos con un alto grado de fiabilidad. En el plano de la vigilancia mono-sensorial, y más concretamente, la visual, son ya varios los resultados publicados que se basan en el empleo de sistemas multi-agente. En vigilancia de tráfico rodado, Monitorix [5] es un sistema multi-agente totalmente descentralizado bajo una plataforma y lenguaje de comunicación de agentes FIPA. El seguimiento de los automóviles se realiza por medio de un modelo de tráfico y de unos algoritmos de aprendizaje que van ajustando los parámetros del modelo. El equipo VSAM (Video Surveillance and Monitoring) ha desarrollado un sistema de vigilancia multi-cámara que permite que un único operador humano monitorice las actividades a partir de un conjunto de sensores de vídeo activos [6]. El sistema permite detectar automáticamente personas y vehículos y tenerlos localizados respecto de un modelo geo-espacial. Otro trabajo reciente propone una arquitectura multi-agente para la comprensión de las dinámicas de una escena por medio de la unión de la información capturada desde diversas cámaras [7]. Si nos adentramos en sistemas multi-agente para vigilancia multi-sensorial nos encontramos con los interesantes trabajos de Molina y otros [8, 9]. Describen un esquema de gestión mediante lógica borrosa para la evaluación de prioridades de tareas multisensoriales en aplicaciones de vigilancia en defensa, todo ello soportado bajo un sistema multi-agente para la lógica de razonamiento.

A continuación se presenta el análisis y diseño de un SMA que realiza el control de un sistema inteligente de vigilancia multi-sensorial, cuyo desarrollo sigue la metodología INGENIAS. En la sección 2 se presentan los conceptos básicos de INGENIAS que se utilizarán posteriormente en este trabajo. En la sección 3 se analizan los requisitos del sistema con el propósito de determinar los objetivos que debe satisfacer el SMA. La descomposición de estos objetivos permitirá determinar un conjunto de tareas y flujos de trabajo en el sistema, así como los agentes responsables de los mismos, determinando la arquitectura del SMA que se presenta en la sección4. Finalmente, en la sección 5 se resumen las principales características del sistema desarrollado así como aquellos aspectos que se han identificado para una evolución del mismo.

2 La metodología INGENIAS

Dentro de las distintas propuestas metodológicas para la construcción de sistemas multi-agente (SMA), INGENIAS [10] es interesante por integrar la mayor parte de los conceptos existentes en teoría de agentes y estar orientada no sólo al análisis, sino también al diseño e implementación de aplicaciones reales. Para ello, proporciona un conjunto de herramientas, el INGENIAS Development Kit (IDK), que facilita el modelado e implementación de SMA:

- Un editor gráfico para modelar los sistemas multi-agente. Este editor permite utilizar el lenguaje INGENIAS o una notación similar a UML. Además este editor se puede personalizar para un dominio de aplicación concreto. Esto permitiría crear, por ejemplo, un editor especializado para el ámbito de los sistemas de vigilancia.
- Módulos de generación de código. Estos módulos permitirán transformar el modelo gráfico del SMA en un conjunto de programas ejecutables, los agentes (sobre la plataforma de ejecución final deseada), salvando así la distancia entre el modelado y la programación.
- Módulo de generación de documentación. Similar a un módulo de generación de código pero lo que genera es un conjunto de páginas HTML que permiten documentar un modelo de sistema social.
- Módulos de verificación de propiedades. Es posible analizar si un modelo cumple un conjunto de requisitos. Para ello, estos módulos recorren el modelo analizando la satisfacción de las propiedades para los que hayan sido diseñados.

El lenguaje de modelado INGENIAS está estructurado en cinco paquetes, que representan los puntos de vista que se pueden considerar para definir un SMA: organización, agentes, objetivos/tareas, interacciones y entorno.

La organización del sistema multi-agente determina el marco en el que los agentes conviven. Define relaciones estructurales (grupos de agentes, jerarquías), normas sociales (limitaciones y formas en el comportamiento de los agentes y sus interacciones), y procesos (en inglés, *workflows*, que determinan cómo colaboran los agentes realizando tareas de la organización). Una organización se estructura en *grupos*. Puede haber varias formas de estructurar una organización. Por ejemplo, de acuerdo a necesidades funcionales. O al mismo tiempo también se podría considerar otra estructuración por distribución geográfica. Un agente, por tanto, puede pertenecer en un momento dado a varios grupos. En general, para dar más flexibilidad a la definición de organizaciones se utiliza el concepto de *rol*, que representa un conjunto de funcionalidad o servicios en una estructura organizativa. Los agentes juegan roles en la organización. Varios agentes pueden jugar el mismo rol, cada uno de forma distinta atendiendo a sus capacidades y estrategias. En cuanto a los procesos, reflejan la dinámica de la organización. Un proceso está definido por un conjunto de tareas o actividades que fluyen a través de la organización (de ahí la denominación inglesa de *workflow*). Las tareas en un proceso producen resultados que pueden ser utilizados por otras para producir nuevos resultados. Las tareas, asimismo, serán ejecutadas por agentes que requerirán para ello de recursos de la organización.

El comportamiento de los agentes viene determinado por su estado mental. El es-

tado mental es el conjunto de objetivos y creencias que tiene el agente en un momento dado. Además, el agente tiene un procesador de estado mental que le permite decidir qué tarea realizar y un gestor de estado mental para crear, modificar o eliminar elementos del estado mental. INGENIAS no explicita cómo se define el procesador de estado mental porque se considera que hay formas muy variadas de realizarlo. Por ejemplo, podría ser un motor de inferencia sobre un conjunto de reglas, razonamiento basado en casos, o una red neuronal. Dependerá de las necesidades de la aplicación o el mecanismo más adecuado según el desarrollador.

Los agentes son entidades intencionales, esto es, actúan porque persiguen unos objetivos. Como además son entidades sociales, colaboran para conseguir satisfacer objetivos de la organización. A la hora de diseñar un SMA se puede empezar identificando objetivos de la organización (del sistema) y descomponerlo en otros más sencillos sucesivamente hasta llegar a objetivos más concretos para los cuales se puedan definir tareas específicas que puedan conducir a su satisfacción. Otra posibilidad es identificar objetivos individuales para los agentes, que también podrían descomponerse de manera similar. En ambos casos, al final habrá una relación entre objetivos y tareas.

Como entidades sociales, los agentes interactúan entre sí. Las interacciones se pueden producir de muchas maneras, siendo las más comunes el intercambio de mensajes o la utilización de espacios comunes donde los agentes pueden actuar (produciendo modificaciones) y percibir (un ejemplo de este segundo caso es una pizarra compartida). Además, y a diferencia de la mayoría de las metodologías orientadas a agentes, en INGENIAS otro aspecto fundamental es la intencionalidad de la interacción: qué objetivos persiguen las partes en una interacción.

El entorno es lo que los agentes perciben y donde pueden actuar. Dependiendo de la aplicación, la percepción y actuación tienen significados muy variados. El entorno estará constituido por un conjunto de recursos, aplicaciones y otros agentes. En muchas ocasiones el entorno se puede especificar como un conjunto de interfaces de aplicación, que serían las clases que lo recubren o que permiten interactuar con él. De hecho, si el entorno son librerías u otras aplicaciones.

Dependiendo del conocimiento que se tenga del sistema o de su naturaleza, el proceso de desarrollo estará guiado por alguno de estos puntos de vista. Por ejemplo, si se pueden establecer fácilmente los requisitos como casos de uso, a partir de éstos se pueden identificar los objetivos y dirigir el análisis y diseño por la descomposición de los mismos, identificando tareas y agentes responsables. O si se conocen los procesos (workflows) de una organización, se definirán las tareas, los agentes y sus responsabilidades en torno a estos procesos.

3 Análisis de los requisitos y objetivos del sistema

El sistema inteligente de vigilancia multi-sensorial tiene que considerar la diversidad de elementos que lo constituyen así como los requisitos por parte del usuario. En particular, en un sistema multi-sensorial se integran sensores que capturan distinto tipo de información y que pueden manipularse diferentemente. Los sensores pueden ser *activos*, si pueden procesar la información que captan, clasificarla y decidir qué

hacer en función del dictamen, o *pasivos*, en cuyo caso mandan la información a algún agente gestor de sensores pasivos que la procesa y toma las decisiones respectivas. Para el tratamiento de la señal de los sensores, en el caso de datos visuales y sonoros, el software debe ser capaz de codificarlos en las formas más simples posibles; la comprensión de escenas o extractos de audio implica la extracción de información semántica, para ir del análisis de bajo nivel a la interpretación automática de alto nivel. Debido a la naturaleza ruidosa e impredecible de las señales de video y audio que son captadas, generalmente es necesario el uso extensivo de la teoría de la probabilidad [11] (redes bayesianas, cadenas ocultas de Markov) o de técnicas basadas en la permanencia espacio-temporal de las señales [12, 13] para detectar y segmentar, así como para almacenar información, analizarla e intercambiarla.

Es evidente que gran parte del esfuerzo empleado en la inteligencia del sistema depende del tipo de información que proporcionan los sensores, un detector de humos no proporciona tantos datos como una cámara, por lo tanto la inteligencia empleada para esta última debe ser más elaborada.

Además, es posible controlar el funcionamiento de los sensores. Por ejemplo, sensores direccionales se puede modificar la orientación de sus coordenadas en altura y anchura (dos ejes de movimiento) y también en profundidad (aplicar zoom o ganancia), con operaciones del tipo:

- *Zoom-In*, *zoom-Out*. Controla la profundidad del zoom en las cámaras y de la ganancia en los micrófonos.
- *Move-Up*, *move-Down*, *move-Left* y *move-Right*. Controla el movimiento del sensor en el eje vertical y horizontal.

Para sensores instalados en vehículos motorizados habrá que decidir además los estados en el movimiento de la plataforma:

- *Drive-Straight*, *drive-Left* y *drive-Right*. La dirección que tomará la plataforma, recto, a izquierda y a derecha, en el caso de una plataforma que se mueve en un plano horizontal, para una plataforma aérea habría que especificar también si sube o baja.
- *Engine-Ahead*, *engine-Back* y *engine-Stop*. Para especificar el sentido de avance de la plataforma, aplicar el freno y poder maniobrar.

El movimiento de los sensores implica que debe existir un *mapa*, que indique por dónde pueden desplazarse los sensores.

Tal como se refleja en la Fig. 1, el usuario humano de este sistema, el *vigilante*, podrá gestionar el mapa para definir el entorno a vigilar, solicitar la inspección de áreas concretas, cotejar alarmas y realizar seguimiento de objetos móviles.

El análisis a partir de los casos de uso puede proceder de varias maneras, dependiendo del conocimiento que se tenga del sistema a realizar. Una posibilidad es asociar un objetivo global del SMA a cada caso de uso y analizar su descomposición en sub-objetivos hasta llegar a un un nivel de concreción a partir del cual se pueden determinar tareas que lleven a la satisfacción de esos objetivos. Obsérvese que para realizar un objetivo hay varias estrategias posibles, esto es, varios planes de tareas posibles. Este tipo de análisis es bastante adecuado cuando se tiene claro el propósito del sistema pero no tanto la forma de llevarlo a cabo o porque hay muchas alternativas o estrategias posibles.

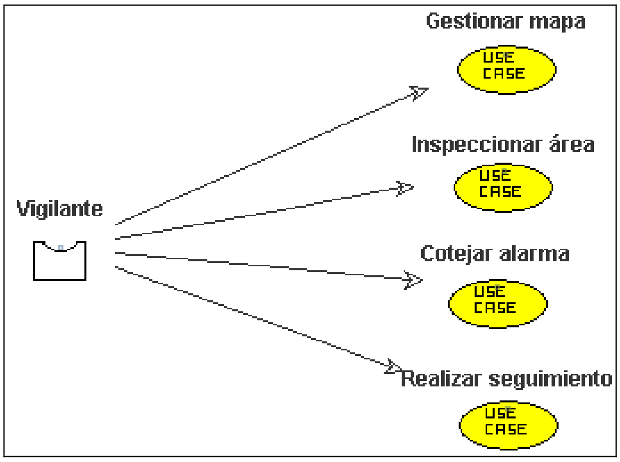


Fig. 1. Diagrama de casos de uso que refleja los requisitos funcionales del usuario

Otra posibilidad es partir de los procesos (workflows) de la organización, si son conocidos, como es el caso en los sistemas de vigilancia, donde habitualmente están bien establecidos los procedimientos de actuación. En este caso el SMA lo que ofrece son facilidades para realizar dichos procesos. En este caso el análisis entra en lo que es diseño ya que parte de los requisitos son precisamente los flujos de actividad del sistema que están regulados previamente. Cada caso de uso determinará uno o varios workflows que definen relaciones entre tareas, que se plasmarán en interacciones entre agentes. Para cada tarea se indicarán los responsables (roles o agentes), los recursos que requieren para su ejecución, sus entradas y productos (salidas). Todos estos aspectos se desarrollan en la siguiente sección.

4 Arquitectura del sistema multi-agente

La estructura organizativa se puede realizar en varios planos. En la Fig. 2 se identifican varios grupos de agentes teniendo en cuenta la funcionalidad. En cada grupo hay varios tipos de agentes especializados de acuerdo a los tipos de dispositivos o funciones concretas del sistema. Así se consideran los sensores, que pueden estar agrupados de acuerdo a sus capacidades, y un grupo de agentes de coordinación que están especializados en determinadas funciones como realizar seguimiento de elementos móviles (personas o animales) o un agente de correlación para comprobar la consistencia de distintos eventos.

También sería posible definir otras estructuras organizativas, por ejemplo, atendiendo a la localización geográfica de los agentes. Por ejemplo, teniendo en cuenta que la coordinación de sensores se hace a nivel de áreas. Un área es una región de un mapa controlada por un *gestor de áreas*. El *agente de seguimiento* utiliza los servicios de un *gestor de áreas* para monitorizar sucesos en el mapa. Dentro de esta otra visión estructural de la organización los agentes desempeñarán distintos roles, tal como se expresa en la Fig. 3. Asimismo en la figura se muestran otros roles que deben desem-

peñar los agentes. Los roles definen las responsabilidades funcionales de los agentes.

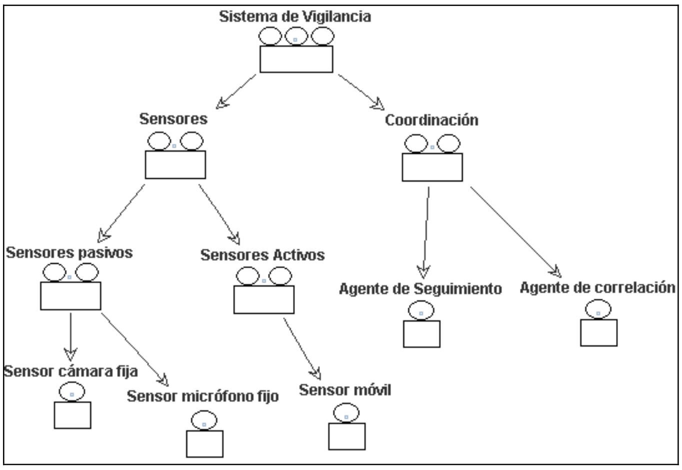


Fig. 2. Estructura organizativa del SMA atendiendo a la funcionalidad

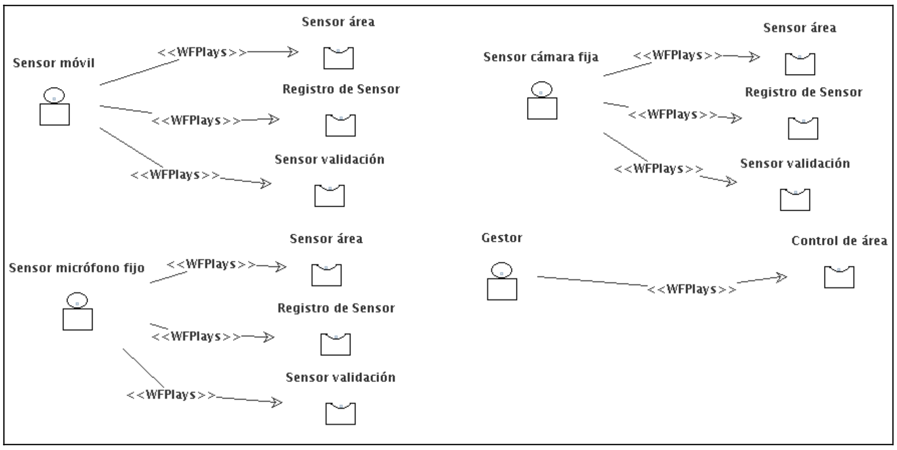


Fig. 3. Roles de los agentes para gestionar la distribución por áreas

Para realizar sus operaciones, algunos agentes necesitan conectarse a uno o varios sensores. Los sensores se representan con los objetos *sensor fijo* y *sensor vehículo*, que permiten enlazar la capa de agentes con elementos externos (véase la Fig. 4). Se asume que los objetos *sensor fijo* y *sensor vehículo* ofrecen un API a los agentes con el que pueden ser notificados cuando ocurre algún evento programado. Al recibir el evento o eventos, los agentes realizan una primera tarea de procesamiento, atendiendo a las necesidades existentes de monitorización.

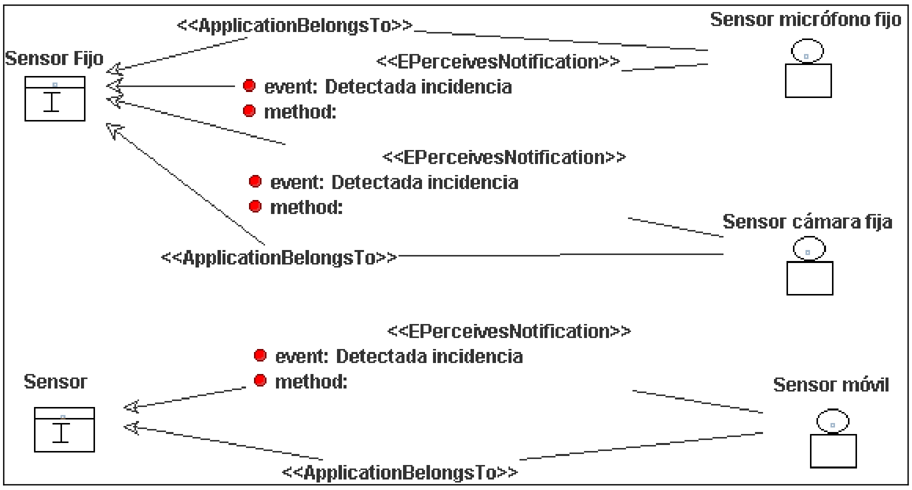


Fig. 4. Utilización de dispositivos físicos por los agentes responsables de sensores

Dada la posibilidad de que en cada momento un área registre un número diferente de sensores, se hace necesario establecer un protocolo de alta y baja en el área. Los sensores fijos harán uso de este protocolo una vez, pero los móviles pueden hacerlo varias veces. De esta forma, se modela la posibilidad de que un agente sensor móvil se desplace de área en área. Este proceso se puede modelar con el flujo de trabajo de la Fig. 5, una representación similar a los *swim-lines* de UML. Aquí se indica que la capacidad de registrarse en un área exige la existencia de dos agentes despempeñando dos roles: el de *control de área* y el de *registro de sensor*. Para indicar la temporización de la ejecución de estas tareas, se utiliza una interacción.

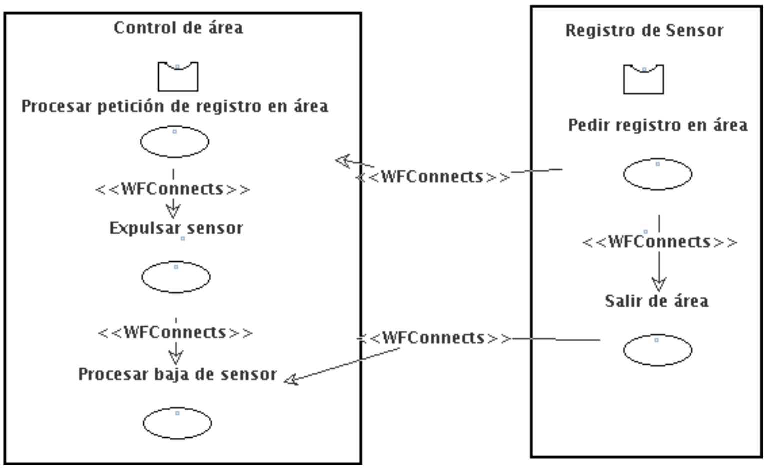


Fig. 5. Workflow para la gestión de altas y bajas en un área

Como indica la figura, inicialmente se solicita el registro con una tarea *Pedir registro en área*. A continuación, el *Control de área* actualiza la definición de su área para tener en cuenta el nuevo sensor. En este caso no tiene sentido que se rechace la peti-

ción, así que se omiten las partes referentes a una posible anulación. En algún momento posterior, el rol *Registro de Sensor* solicitará su baja de un área, baja que es procesada por *Procesar baja de sensor*. De forma alternativa, el controlador de área puede optar por expulsar el sensor. Este sería el caso de un sensor defectuoso que proporciona señales no correspondidas con lo que registran otros. Como la tarea *expulsar sensor* se considera sólo en esta circunstancia, no se vería necesario informar al expulsado. Las interacciones asociadas se definen en diagramas de interacción, como el de la Fig. 6. Los participantes de la interacción son roles desempeñados por agentes de los ya conocidos. El motivo de especificar las interacciones con roles es poder reutilizar capacidades ya especificadas para otros agentes. Asociada a la interacción, se declara su propósito, que es *Monitorizar el área*. Esta asociación sirve para indicar en qué contribuye esta interacción a la funcionalidad del sistema, expresada como los objetivos que persigue.

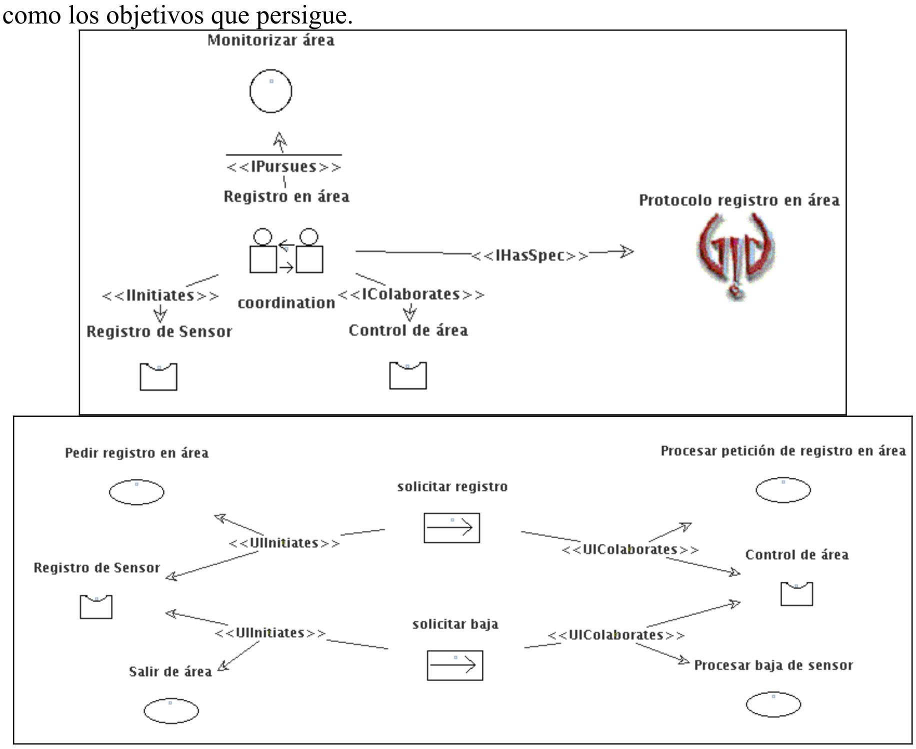


Fig. 6. Interacción para solicitar el registro en un área y protocolo asociado

El protocolo de registro (Fig. 6) requiere la temporización de las cuatro tareas básicas ya vistas. La unidad de interacción *solicitar registro* transfiere la petición elaborada por la tarea *pedir registro en área* al estado mental del agente que desempeña el rol *control de área*. Una vez allá, es procesada por la tarea *procesar petición de registro en línea.* De forma similar, la unidad de interacción *solicitar baja* contacta con el *control de área* para comunicar el resultado de ejecutar *salir de área.*

Refinando el diseño de las interacciones y los flujos de tareas es posible describir el comportamiento de los agentes. Cada agente tiene un estado mental que consta de sus objetivos (responsabilidades asignadas a los agentes en los workflows) y creencias (que pueden darse como resultado de ejecución de tareas por el mismo u otros agentes). Además, se determinarán reglas de transición de los agentes mediante la asociación de estados mentales y tareas en diagramas de agente. Con toda esta información el IDK puede generar código utilizando plantillas para la plataforma donde se vayan a ejecutar cada agente.

5 Conclusiones

Los sistemas inteligentes de vigilancia constan de una gran diversidad de entidades que tienen que cooperar en entornos distribuidos altamente dinámicos. La utilización de agentes para su control permite añadir mayor grado de autonomía y respuesta al sistema gracias a sus capacidades de adaptación y cooperación. Esto añade flexibilidad al diseño pero exige asimismo cierta disciplina metodológica ya que el número de componentes del sistema es grande y sus interrelaciones pueden ser complejas. Por esta razón se ha considerado la utilización de una metodología orientada a agentes, INGENIAS, para su desarrollo. Esencialmente, este tipo de metodologías facilita la definición de los aspectos organizativos, a partir de los cuales se va refinando la funcionalidad, comportamiento e interacciones de los agentes que lo componen. La visión organizacional del sistema también ayuda a integrar procesos de trabajo (workflows) ya conocidos en la gestión de sistemas de vigilancia, con lo cual el diseño del sistema multi-agente es más comprensible para los expertos en sistemas de vigilancia y no solo para el ingeniero informático. Además, la concepción del sistema utilizando agentes facilita la incorporación paulatina de nueva funcionalidad y servicios mediante la inclusión de nuevos agentes, por ejemplo para definir nuevas tareas de coordinación.

Para que el desarrollo metodológico sea efectivo es importante disponer de herramientas de apoyo para su aplicación. En este caso particular es especialmente útil la facilidad de generación de código del INGENIAS Development Kit (IDK), ya que permite la generación de código en varios entornos. De esta manera, es posible realizar un diseño donde se consideren todos los componentes de manera integrada, al mismo nivel de abstracción (independiente de la plataforma), y posteriormente generar código específico para cada componente en su plataforma concreta de ejecución (servidores con Java para los agentes de coordinación y agentes sensores en algunos dispositivos con lenguajes de programación y middleware específico).

La base de meta-modelado del entorno IDK plantea además, como trabajo futuro, la posibilidad de definir un entorno especializado para la construcción de configuraciones de control de sistemas de vigilancia inteligente donde en vez de elementos específicos de la teoría de agentes se diseñe con conceptos de sistemas de vigilancia.

Agradecimientos

Este trabajo ha sido realizado en el contexto de los proyectos "*Métodos y herramientas para modelado de sistemas multiagente"* y *"Diseño e implementación de un sistema de atención visual selectiva para monitorización dinámica y distribuida de escenarios con distintos tipos de objetos móviles y deformables"*, subvencionados por el Ministerio de Educación y Ciencia, referencias TIN2005-08501-C03-01 y TIN2004-C07661-C02-02, respectivamente, así como el proyecto PBI06-0099 "*Un sistema multiagente para el control inteligente de cámaras PTZ fijas y móviles en redes inalámbricas*", subvencionado por la Junta de Comunidades de Castilla-La Mancha.

Referencias

1. Molina, J.M., García, J., Jiménez, F.J. & Casar, J.R. Surveillance multisensor management with fuzzy evaluation of sensor task priorities. *Engineering Applications of Artificial Intelligence 12* (2002) 511-527
2. Conci, N., De Natale, F.G.B., Bustamante, J. & Zangherati, S. A wireless multimedia framework for the management of emergency situations in automotive applications: The AIDER system. *Signal Processing: Image Communication 20* (2005) 907-926
3. Boult, T.E., Micheals, R., Gao, X., Lewis, P., Power, C., Yin, W. & Erkan, A. Frame-rate omnidirectional surveillance & tracking of camouflaged and occluded targets. En: *Second IEEE Workshop on Visual Surveillance* (1999) 48
4. Goradia, A. Cen, Z., Xi, N. & Mutka, M. Modeling and design of mobile surveillance networks using a mutational analysis approach. En: *Proceedings of 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2005* (2005)
5. Abreu, B., et al. A. Video-based multi-agent traffic surveillance system. En: *Proceedings of the IEEE Intelligent Vehicles Symposium, IV2000* (2000) 457-462
6. Collins, R.T., Lipton, A.J., Fujiyoshi, H. & Kanade, T. Algorithms for cooperative multisensor surveillance. *Proceedings of the IEEE 89:10* (2001) 1456 - 1477
7. Remagnino, P., Shihab, A.I., Jones, G.A. Distributed intelligence for multi-camera visual surveillance. *Pattern Recognition 37:4* (2004) 675-689
8. Molina, J.M., García, J., Jiménez, F.J. & Casar, J.R. Fuzzy reasoning in a multi agent system of surveillance sensors to manage cooperatively the sensor-to-task assignment problem. *Applied Artificial Intelligence 18:8* (2004) 673-711
9. Molina, J.M., García, J., Jiménez, F.J. & Casar, J.R. Cooperative management in a net of intelligent surveillance agent-sensors. *Int. Journal of Intelligent Systems 18:3* (2003) 279-307
10. Pavón, J., Gómez-Sanz, J.J. & Fuentes, R. The INGENIAS methodology and tools. En: Henderson-Sellers, B. and Giorgini, P., editors: *Agent-Oriented Methodologies*. Idea Group Publishing (2005), 236-276
11. Oliver, N.M., Rosario, B. & Pentland, A.P. A bayesian computer vision system for modeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22:8* (2000) 831-843
12. López, M.T., Fernández-Caballero, A., Fernández, M.A., Mira, J. & Delgado, A.E. Motion features to enhance scene segmentation in active visual attention. *Pattern Recognition Letters 27:5* (2006) 469-478
13. Fernández-Caballero, A., Fernández, M.A., Mira, J. & Delgado, A.E. Spatio-temporal shape building from image sequences using lateral interaction in accumulative computation. *Pattern Recognition 36:5* (2003) 1131-1142

Simulación de sistemas sociales con agentes software

Juan Pavón[1], Millán Arroyo[2], Samer Hassan[1] y Candelaria Sansores[1]

[1] Universidad Complutense Madrid, Facultad de Informática, Ciudad Universitaria s/n, 28040 Madrid, España
jpavon@sip.ucm.es, samer2004@gmail.com, csansores@fdi.ucm.es
[2] Universidad Complutense Madrid, Facultad de Ciencias Políticas y Sociología, Dep. Sociología IV, Ciudad Universitaria s/n, 28040 Madrid, España
millan@cps.ucm.es

Resumen. La simulación con agentes software abre nuevas posibilidades para el estudio de fenómenos sociales. La teoría de agentes software facilita el modelado de los aspectos organizativos y de comportamiento de los individuos de una sociedad. Un agente puede representar un individuo en una sociedad, que percibe y reacciona ante los eventos de su entorno de acuerdo a su estado mental (creencias, deseos, intenciones), y que interacciona con otros agentes de su entorno social. Existen herramientas para realizar la simulación con agentes pero éstas requieren un conocimiento de técnicas de programación que normalmente son ajenas a los sociólogos. Para salvar esta dificultad proponemos la utilización de lenguajes gráficos de modelado de agentes adaptados al domino de estudio sociológico junto con herramientas de generación de código ejecutable en las plataformas de simulación. La construcción de este marco de trabajo está basada en los métodos y herramientas de INGENIAS, una metodología para el desarrollo de sistemas multi-agente.

1 Introducción

Los fenómenos sociales son altamente complejos, pues se hallan imbricados en complejas redes de interacción e interdependencia mutua. Las explicaciones sociológicas requieren asimismo de modelos explicativos complejos, en los que múltiples factores (cambiantes) interactúan con el fenómeno que se trata de explicar y entre sí mismos. La simplificación de los modelos explicativos, en parte deseable, es especialmente delicada, pues a menudo dan lugar a artefactos teóricos poco fieles con las observaciones empíricas.

Un sistema social está constituido por un colectivo de individuos que interactúan mutuamente entre sí o a través de artefactos de su entorno social. Estos individuos evolucionan de forma autónoma (tienen su idiosincrasia particular), están motivados por sus propias creencias y objetivos personales, y las circunstancias del entorno social en el que se mueven moldean en gran medida dichas creencias y objetivos personales. No se nos debe escapar que esas creencias y objetivos pueden a su vez evolucionar en el tiempo para cada individuo, bien sea por su peculiar peripecia biográfica, irrepetible, o por el momento que atraviesa en su ciclo vital o por los cambios

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 389-400, 2006.

de las tendencias sociales (cambios en el entorno). En relación con esto último, es importante señalar que la población evoluciona demográficamente y esto tiene repercusiones a nivel micro y a nivel macro. A nivel micro, los individuos están sujetos probabilísticamente a pautas de *ciclo de vida* (marcadas fundamentalmente por la demografía, aunque no solo): se emparejan, se reproducen y finalmente mueren, pasando por diversas etapas en las que se sujetan a distintos patrones de comportamientos e intenciones. Dichas tendencias demográficas a su vez, a nivel macro, tienen implicaciones sobre el sistema social (influyen en éste y son influidas por éste) y por tanto capacidad de interacción dinámica con otros procesos sociales. Por otro lado, es preciso tener en cuenta que todos los fenómenos sociales son absolutamente contingentes, y por tanto impredecibles y cambiantes. No están sujetos a leyes, sino a tendencias, las cuales pueden afectar a los individuos probabilísticamente. La indeterminación de los procesos y los sistemas sociales es mucho mayor que en los sistemas físicos o incluso los biológicos.

Todos estos determinantes contribuyen a que un sistema social sea altamente dinámico y complejo y por ello su reductibilidad mediante el mero recurso a la modelización matemática (modelización mediante ecuaciones estructurales, análisis estadísticos multivariantes o tratamientos estadísticos de series temporales) adolece de serias limitaciones, tanto desde el interés explicativo como predictivo. El principal problema de la modelización matemática es que sólo permite jugar con tendencias (probablilísticas) y no considera el comportamiento original del sujeto, pieza básica de cualquier sistema social. Los efectos dinámicos de procesos altamente retroalimentados no quedan bien reflejados en los tratamientos matemáticos-estadísticos, y sin embargo dichos procesos son consustanciales a los sistemas sociales reales.

Por su parte, un sistema multi-agente (SMA) consta de un conjunto de entidades software autónomas (los agentes) que interaccionan mutuamente y con su entorno. El hecho de ser autónomos significa que los agentes son entidades activas que pueden tomar sus propias decisiones. Esto no es así, por ejemplo, con objetos, que están predeterminados a realizar las operaciones que se les soliciten. Un agente, sin embargo, decidirá si realiza o no una operación solicitada, para lo cual tendrá en cuenta sus objetivos y prioridades, así como el contexto en que crea encontrarse.

Puede entenderse, por tanto, que el paradigma de agentes se asimila bastante bien a lo que es un sistema social. De hecho, en teoría de agentes software existen numerosos trabajos sobre los aspectos organizativos de los sistemas multi-agente. También se aplican teorías provenientes del campo de la psicología, siendo la más extendida el modelo de Deseos-Creencias-Intenciones (en inglés, *Believes-Desires-Intentions*, BDI) [1].

En este contexto se han desarrollado varias herramientas que permiten la simulación de sistemas complejos, entre ellos los sistemas sociales, utilizando como base el paradigma de agentes software. Un sistema de simulación basada en agentes permite ejecutar un conjunto de agentes, que pueden ser de distintos tipos, en un entorno en el cual se pueden realizar observaciones de su comportamiento. Estas observaciones permiten analizar el comportamiento colectivo emergente y las tendencias de la evolución del sistema. De esta manera se pueden realizar estudios empíricos de los sistemas sociales. Dado que la simulación se realiza en un entorno controlado, en una o varias computadoras, este tipo de herramientas permiten realizar experimentos y estu-

dios que de otra forma serían inviables.

Existen, a nuestro modo de ver, algunas limitaciones que no podemos obviar a la hora de simular sistemas sociales reales. La principal consiste en que el individuo (a diferencia del agente software) es en sí mismo una estructura compleja y por tanto su comportamiento es más imprevisible y menos determinado que el del agente, cuyos comportamientos y capacidades perceptuales se diseñan con relativa sencillez. Además, no es posible en la práctica considerar en la simulación los innumerables matices que cabe encontrar en un sistema social real, en lo que se refiere a la interacción de los agentes, la caracterización del entorno, etc. Por esto ha de entenderse que es imposible o poco práctico pretender la simulación de un sistema social en su globalidad. En cambio, debemos y podemos limitarnos a simular un proceso social concreto en un contexto sistémico e interactivo. Por tanto, la simulación de sistemas sociales debe ser concebida en términos del centramiento sobre un proceso concreto.

Pese a las limitaciones, el paradigma multi-agentes parece adaptarse de forma adecuada a la naturaleza y peculiaridades de los fenómenos sociales, permitiendo superar las limitaciones de los métodos de modelado matemáticos. Sin embargo, los sociólogos y científicos sociales potencialmente interesados en utilizar esta nueva metodología se enfrentan a una dificultad de orden práctico que no debe ignorarse. La utilización de estos sistemas no es sencilla ya que los modelos se especifican como programas, generalmente utilizando un lenguaje de programación orientado a objetos. Esto hace que la definición de modelos por parte de sociólogos no sea una tarea sencilla para éstos, ya que habitualmente no disponen de la adecuada formación informática para esta tarea. Algunos esfuerzos se están realizado para tratar de facilitar el modelado gráfico de estos sistemas, como en Sesam (www.simsesam.de), que permite modelar gráficamente el comportamiento como máquinas de estados finitos usando una librería de comportamientos básicos, o Repast Py (repast.sourceforge.net/repastpy), aunque en este último al final hay que acabar escribiendo scripts con Python. El problema en estas soluciones es que todavía siguen siendo necesarios los conocimientos de programación y que los tipos de sistemas que se pueden modelar son bastante simples (se trata más bien de herramientas de prototipado rápido).

Dentro de la ingeniería de software orientada a agentes existen, sin embargo, lenguajes gráficos de modelado de SMA de mayor nivel de abstracción. Los conceptos utilizados en estos lenguajes pueden ser más cercanos a los que utilizaría un sociólogo y por ello nos parece que pueden ser apropiados para resolver el problema que hemos planteado. Partiendo de esta hipótesis, proponemos la utilización de una metodología de desarrollo de SMA, concretamente, INGENIAS [1]. Esta elección se debe a que INGENIAS proporciona un conjunto de herramientas, el INGENIAS Development Kit (IDK), que facilitará su aplicación para modelar, y posteriormente simular, sistemas sociales:

- Un editor gráfico para modelar los sistemas multi-agente. Este editor permite utilizar el lenguaje INGENIAS o una notación similar a UML. Pero lo más interesante para la simulación social es que el editor se puede personalizar para un dominio de aplicación concreto (se ha realizado para sistemas holónicos en la Univ. Politécnica de Valencia [1], por ejemplo). Esto permitirá crear editores especializados para ámbitos de estudios sociológicos específicos.
- Módulos de generación de código. Estos módulos permitirán transformar el mo-

delo gráfico en un programa ejecutable en un entorno de simulación, salvando así la distancia entre el modelado y la programación. Además, sería posible generar código para varios entornos de simulación basada en agentes, lo cual es interesante para replicar el mismo modelo sobre distintos entornos y validar mejor los resultados obtenidos.

- Módulos de verificación de propiedades. Es posible analizar si un modelo cumple un conjunto de requisitos. Para ello, estos módulos recorren el modelo analizando la satisfacción de las propiedades para los que hayan sido diseñados.
- Módulo de generación de documentación. Similar a un módulo de generación de código pero lo que genera es un conjunto de páginas HTML que permiten documentar un modelo de sistema social.

En la siguiente sección se presentan elementos del lenguaje de modelado INGENIAS para SMA que pueden ser útiles para el modelado de sistemas sociales. Como para el estudio de estos últimos se requieren algunas facilidades que no se habían tenido en cuenta, por no ser inicialmente necesarias para el desarrollo de SMA, ha sido necesario considerar extensiones, que se presentan en la sección 3. Para ilustrar cómo sería el modelado de uno de estos sistemas, en la sección 4 se muestra un caso real desarrollado en colaboración con profesores de la Facultad de Sociología. Para concluir, en la sección 5 se discuten las principales contribuciones de este trabajo y los aspectos que se plantean para su mejora.

2 El lenguaje de modelado INGENIAS

INGENIAS es una metodología para la construcción de sistemas multi-agente (SMA) que integra gran parte de propuestas que se han desarrollado en este ámbito. Esta integración se ha producido mediante la experimentación en la realización de múltiples aplicaciones de agentes a lo largo de los últimos años. Por esta razón, y como proyecto de investigación, INGENIAS asume desde el principio la evolución del lenguaje de modelado para SMA que utiliza en sus métodos y herramientas. Para facilitar esta evolución la especificación del lenguaje de modelado INGENIAS está basada en un lenguaje de meta-modelado, concretamente MOF (*Meta-Object Facility*) [5], un estándar del OMG (*Object Management Group*).

Las herramientas, el INGENIAS Development Kit (IDK), están generadas a partir de la especificación (meta-modelo) del lenguaje de modelado INGENIAS. De esta manera, si se cambia el meta-modelo, por ejemplo para añadir o refinar algún concepto, se puede generar automáticamente una nueva versión del IDK para tenerlo en cuenta. De la misma manera es posible personalizar las herramientas para un lenguaje específico simplemente describiendo las extensiones que requiera sobre el meta-modelo de INGENIAS. Así la evolución de INGENIAS resulta relativamente sencilla.

El lenguaje de modelado INGENIAS está estructurado en cinco paquetes, que representan los puntos de vista que se pueden considerar para definir un SMA (véase la Fig. 1): organización, agente, objetivos/tareas, interacciones y entorno. A continuación se introducen brevemente los elementos más relevantes de cada uno de estos puntos de vista. Su utilización y la notación gráfica asociada se ilustran en el ejemplo

de la sección 4.

La **organización** del sistema multi-agente determina el marco en el que los agentes conviven. Define relaciones estructurales (grupos de agentes, jerarquías), normas sociales (limitaciones y formas en el comportamiento de los agentes y sus interacciones), y procesos (en inglés, *workflows*, que determinan cómo colaboran los agentes realizando tareas de la organización). Una organización se estructura en *grupos*. Puede haber varias formas de estructurar una organización. Por ejemplo, de acuerdo a necesidades funcionales. O al mismo tiempo también se podría considerar otra estructuración por distribución geográfica. Un agente, por tanto, puede pertenecer en un momento dado a varios grupos. En general, para dar más flexibilidad a la definición de organizaciones se utiliza el concepto de *rol*, que representa un conjunto de funcionalidad o servicios en una estructura organizativa. Los agentes juegan roles en la organización. Varios agentes pueden jugar el mismo rol, cada uno de forma distinta atendiendo a sus capacidades y estrategias. En cuanto a los procesos, reflejan la dinámica de la organización. Un proceso está definido por un conjunto de tareas o actividades que fluyen a través de la organización (de ahí la denominación inglesa de *workflow*). Las tareas en un proceso producen resultados que pueden ser utilizados por otras para producir nuevos resultados. Las tareas, asimismo, serán ejecutadas por agentes que requerirán para ello de recursos de la organización. Ambos aspectos, estructural y dinámico, definen la visión *macro* del sistema multi-agente. Esta perspectiva facilita la gestión de sistemas complejos ya que permite determinar el contexto y normas de actuación de los agentes, al igual que ocurre cuando se trata de organizaciones humanas.

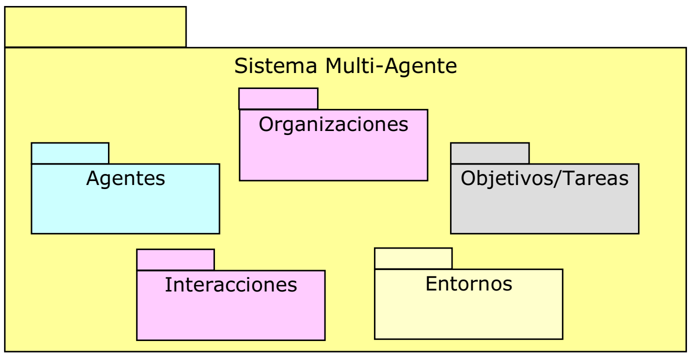


Fig. 1. Puntos de vista de un SMA según INGENIAS

El comportamiento de los **agentes** viene determinado por su estado mental. El estado mental es el conjunto de objetivos y creencias que tiene el agente en un momento dado. Además, el agente tiene un procesador de estado mental que le permite decidir qué tarea realizar y un gestor de estado mental para crear, modificar o eliminar elementos del estado mental. INGENIAS no explicita cómo se define el procesador de estado mental porque se considera que hay formas muy variadas de realizarlo. Por ejemplo, podría ser un motor de inferencia sobre un conjunto de reglas, razonamiento basado en casos, o una red neuronal. Dependerá de las necesidades de la aplicación o el mecanismo más adecuado según el desarrollador.

Los agentes son entidades intencionales, esto es, actúan porque persiguen unos **objetivos**. Como además son entidades sociales, colaboran para conseguir satisfacer objetivos de la organización. A la hora de diseñar un SMA se puede empezar identificando objetivos de la organización (del sistema) y descomponerlo en otros más sencillos sucesivamente hasta llegar a objetivos más concretos para los cuales se puedan definir **tareas** específicas que puedan conducir a su satisfacción. Otra posibilidad es identificar objetivos individuales para los agentes, que también podrían descomponerse de manera similar. En ambos casos, al final habrá una relación entre objetivos y tareas.

Como entidades sociales, los agentes interactúan entre sí. Las **interacciones** se pueden producir de muchas maneras, siendo las más comunes el intercambio de mensajes o la utilización de espacios comunes donde los agentes pueden actuar (produciendo modificaciones) y percibir (un ejemplo de este segundo caso es una pizarra compartida). Además, y a diferencia de la mayoría de las metodologías orientadas a agentes, en INGENIAS otro aspecto fundamental es la intencionalidad de la interacción: qué objetivos persiguen las partes en una interacción.

Finalmente, el **entorno** es lo que los agentes perciben y donde pueden actuar. Dependiendo de la aplicación, la percepción y actuación tienen significados muy variados. El entorno estará constituido por un conjunto de recursos, aplicaciones y otros agentes. En muchas ocasiones el entorno se puede especificar como un conjunto de interfaces de aplicación, que serían las clases que lo recubren o que permiten interactuar con él. De hecho, si el entorno son librerías u otras aplicaciones. Para simulación social, el entorno de los agentes (individuos de una sociedad en tal caso) requiere considerar la localización de los mismos. Este aspecto se trata en la siguiente sección.

3 Extensión de INGENIAS para simulación social

Hay algunos aspectos relativos a la definición de modelos listos para simular que son difíciles de expresar con el lenguaje de modelado INGENIAS en su estado actual. Es por eso que se han planteado extensiones a dicho lenguaje. Esencialmente son dos los aspectos a considerar relativos a las perspectivas espacial y temporal de las simulaciones. Estos aspectos se podrían considerar como extensiones del paquete de entorno.

La perspectiva temporal trata el flujo de tiempo en el modelo durante la ejecución de la simulación. En nuestro caso asumimos que las simulaciones serán dirigidas por tiempo (en vez de por eventos discretos) ya que la mayoría de los entornos de simulación basados en agentes así lo hacen. Por tanto, hace falta modelar pasos de tiempo constantes para simular el ciclo percepción-reacción de los agentes que actúan con el paso del tiempo.

La perspectiva espacial describe los aspectos relacionados con el posicionamiento de los agentes en un espacio. En general, los entornos de simulación basada en agentes proporcionan espacios de dos y tres dimensiones con configuraciones muy diversas.

Estas extensiones requieren modificar el meta-modelo de INGENIAS y regenerar las herramientas. Así, se puede definir un nuevo entorno de desarrollo IDK especiali-

zado para simulación. Tal como se muestra en la Fig. 2, el sociólogo, como experto del dominio y modelador, definirá y modificará los modelos sociales a simular con el editor del IDK extendido. Desde este editor se pueden invocar distintos tipos de módulos. Normalmente, antes de generar el código para el simulador habrá que verificar que los modelos son correctos de acuerdo a ciertas propiedades. Por ejemplo, que todos los elementos necesarios para la simulación hayan sido definidos y que no haya agentes sin tareas asignadas o aislados completamente. Otro tipo de propiedades que se vayan considerando útiles también se podrán verificar, pero para ello habrá que crear nuevos módulos. Una vez que los modelos cumplen las propiedades requeridas se puede generar el código para un simulador particular. Para ello se invoca el módulo de generación de código correspondiente. El código que se obtiene es fuente y habrá que compilarlo junto con las librerías (paquetes) del simulador. A partir de ese momento se puede utilizar el simulador y sus herramientas para obtener e interpretar los resultados de la simulación.

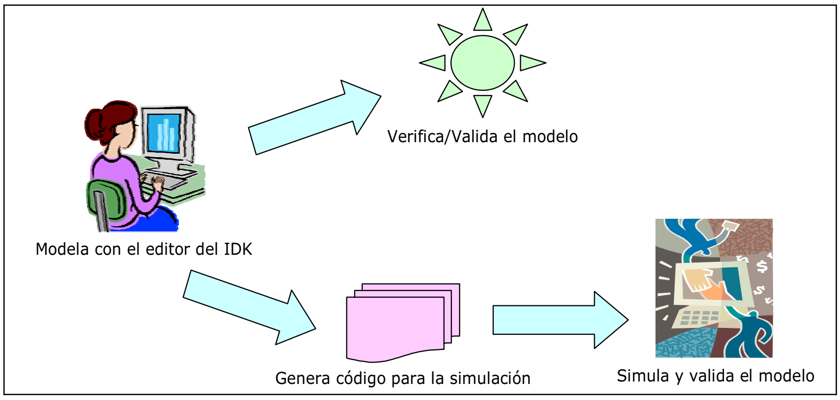


Fig. 2. Desarrollo de modelos de simulación social con el IDK

4 Ejemplo: Estudio de la religiosidad en la sociedad española

Simular la evolución de la religiosidad de los españoles es un caso idóneo para validar nuestra propuesta por tratarse de un proceso social complejo, en la medida que tiene elementos en común con muchos otros problemas susceptibles de ser abordados por sociólogos mediante simulación, especialmente aquellos relacionados con las dinámicas de cambio.

La validación del modelo se realiza comprobando que, a partir de las condiciones iniciales observadas en 1990, la simulación evoluciona ajustándose a los datos empíricos conocidos [1]. Ello permitiría poder realizar predicciones fiables de la evolución social y alcanzar un conocimiento teórico más profundo de cómo y por qué se están transformando los posicionamientos religiosos de los españoles.

Para ilustrar cómo modelar este tipo de problemas con INGENIAS a continuación se muestran algunos aspectos relevantes a partir de un estudio sociológico real sobre

el tema [2]. Es importante en esta discusión observar cómo los conceptos de agentes pueden corresponderse directamente con la terminología utilizada en el discurso sociológico. En España se está experimentando un proceso de secularización intenso caracterizado por un brusco y rápido descenso de la práctica religiosa y de la confianza y credibilidad de la población hacia la institución eclesial, mientras que otros indicadores de otras dimensiones de la religiosidad disminuyen en el tiempo con mucha mayor suavidad que los mencionados (por ejemplo, las creencias, la importancia atribuida a Dios en la vida, o el deseo de espiritualidad, entre otras). Este descenso, además de favorecer la emergencia de contingentes de población no creyente, está favoreciendo la emergencia de nuevas formas de religiosidad, en detrimento de la religiosidad más ortodoxa y tradicional (los que siguen los preceptos de la Iglesia y van a misa regularmente), distinguiendo al menos dos tipos de formas de religiosidad emergentes. Por un lado, es destacable una religiosidad de *baja intensidad* en la cual las funciones religiosas se reducen a una mínima expresión sin desaparecer (tiende a recurrirse a la religión solo o casi solo en momentos especiales y señalados de la vida; rituales de nacimiento, matrimonio, muerte, momentos de tensión por dificultades, grandes cambios vitales o momentos significados de la vida). Y, por otro lado, es también destacable un grupo significado y cada vez mayor de individuos que, sintiéndose religiosos y con una vida religiosa relativamente importante, manifiestan abiertamente su desencuentro con la Iglesia y tratan de vivir su religiosidad ignorando sus preceptos.

Así, para el objeto de este estudio, se pueden considerar cuatro grupos que reflejan las tendencias sociales, tal como se refleja en el diagrama de organización de la Fig. 3. En ésta se utiliza el concepto de organización de SMA para representar (como un icono rectangular con tres círculos) la sociedad y cada colectivo como un grupo (icono rectangular con dos círculos), esencialmente los siguientes:

- *Eclesiales* (22%, en descenso). Católicos que confían en la Iglesia y asisten a misa semanalmente.
- *Laxos* (23%, estables). Católicos que confían en la iglesia y asisten a misa solo ocasionalmente o nunca.
- *Alternativos* (19%, en aumento). Católicos (en su inmensa mayoría, aunque no todos) que se sienten personas religiosas pero que no confían en la Iglesia y no asisten regularmente a misa.
- *Arreligiosos* (35%, en aumento). No confían en la Iglesia y no se consideran personas religiosas.

En cada categoría podrían establecerse ciertas subcategorías, en las que ahora no vamos a entrar. Una distinción de especial relevancia entre los *eclesiales* puede ser la distinción entre religiosos profesionales (predicadores) y *fieles*. Entre los *alternativos* la principal distinción se da entre los que disienten de la jerarquía eclesial desde dentro (comunidades de base, afines a teologías desautorizadas, etc.) y los que reconstruyen su religiosidad ignorando a la Iglesia: movimientos *New Age*, religiosidad individualizada, etc.). Entre los arreligosos la principal distinción sería entre los indiferentes y los abiertamente no creyentes (agnósticos y ateos).

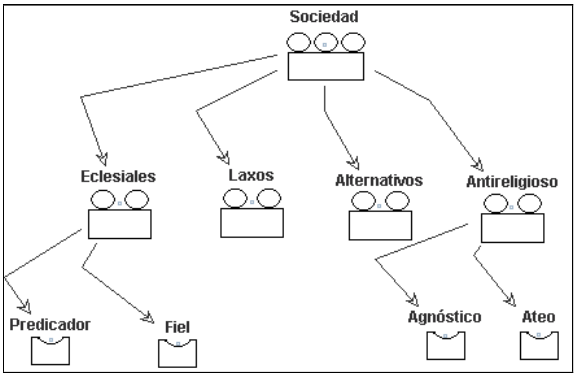


Fig. 3. Grupos sociales en el estudio sobre religiosidad

Otras características a tener en cuenta a la hora de modelar al individuo son las relacionadas con la posición del sujeto en la estructura social: las variables estructurales sociodemográficas y socioeconómicas. Independientemente de que influyan o no en la religiosidad, son necesarias para simular una sociedad, porque definen las pautas de interacción social. Los individuos se comportan y piensan de distintas maneras de acuerdo con dichas características: Sexo, edad, estado civil, si tienen hijos o no y cuantos, estudios, ocupación, y estatus económico, al menos.

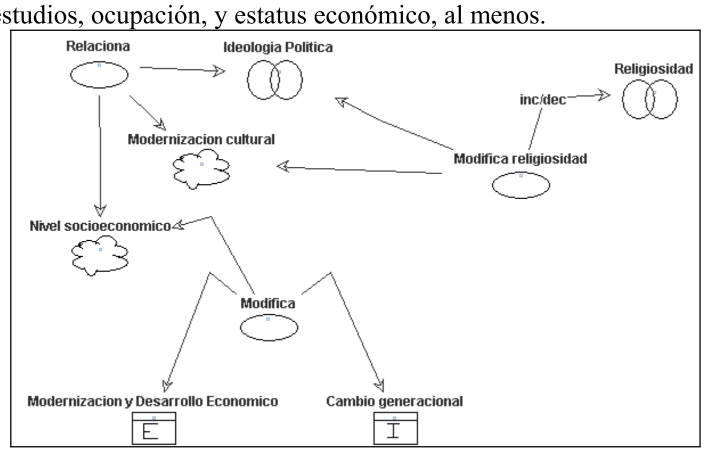


Fig. 4. Ejemplo de proceso de transformación

Estos procesos de transformación de la religiosidad están profundamente imbricados en otros fenómenos sociales con los que interactúan mediante procesos de retroalimentación (positiva o negativa). Se trata de los factores intervinientes que influyen en la pertenencia o cambio de las categorías descritas. Por ejemplo, existen fuertes vínculos entre religiosidad (o irreligiosidad) de un lado y la ideología política y la modernización cultural por otro. Los políticamente conservadores presentan una alta predisposición a mantenerse como eclesiales, mientras que los de izquierda la rechazan fuertemente. Los más modernos tienden a la vez a estar más secularizados (serán mayoritariamente arreligiosos) mientras que los más tradicionales serán mayoritariamente eclesiales.

De la misma manera, la ideología política y la modernización cultural interaccionan mediante retroalimentación entre sí. También la religiosidad mantiene una estrecha relación con otras variables o factores explicativos, como la edad o el género. Con la edad sobre todo como consecuencia del cambio de valores (actitudes, creencias, percepciones y sensibilidades) asociado a la modernización cultural, la cual a su vez incide en las actitudes políticas. Y con el género en la medida que existen importantes diferencias culturales que inciden fuertemente en la orientación religiosa, si bien se atenúan bastante (sin desaparecer) en la media que se asumen los nuevos valores de la modernización cultural.

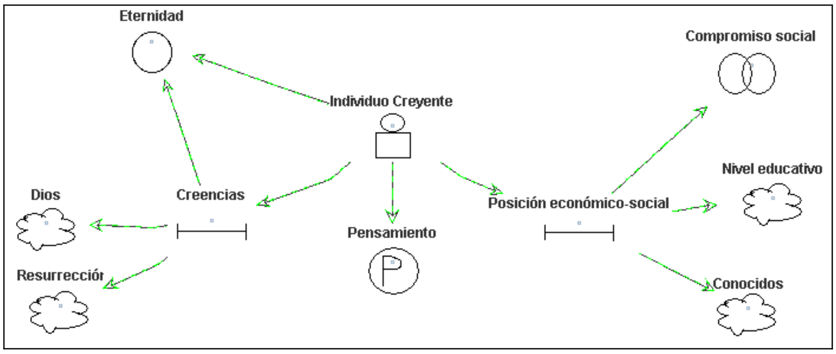


Fig. 5. Modelado del individuo creyente

De otro lado, la modernización cultural no es ajena a la modernización y desarrollo económico, así como a la desigual distribución de la renta, la riqueza y la capacidad adquisitiva. En la medida que los estudios y los contactos interpersonales a alcanzar una posición socioeconómica, contribuirán indirectamente, en alguna medida, a modificar las posiciones políticas y las religiosas. Como también el crecimiento económico incide en el desarrollo de la modernización cultural.

En la tarea de vincular dichos conceptos abstractos a un modelo adaptado a los SMA se debe considerar semánticamente equivalentes el sistema complejo a simular y el conjunto, muy numeroso y descentralizado, de agentes (que representan individuos), situados en un entorno cerrado. Por tanto, *la especificación de características y comportamiento de cada agente se hace esencial*, para que recoja todas las dimensiones que influyan en el problema estudiado, a modo de los índices en la investigación cuantitativa sociológica. Dichas características se transforman en variables interrelacionadas entre sí y con unas reglas de evolución particulares (no varía de igual forma la edad que el nivel económico). En la Fig. 5 se muestra un ejemplo de un agente que representa a un individuo creyente, que tiene como objetivo lograr la eternidad, y como base de creencias conceptos como Dios o la Resurrección. Asimismo tiene un nivel socio-económico constituido por varias variables como su nivel educativo, su red de conocidos (que se asociaría a un grupo en un diagrama de organización, no representado en esta figura), y compromisos adquiridos. La evolución del comportamiento del agente, en este caso los componentes que definen su religiosidad, viene determinado por un procesador de estado mental, que en el modelo se representa como el componente Pensamiento. Este componente será una función o un conjunto de reglas que determinan la evolución de las distintas variables.

Estos agentes-individuos podrán evolucionar dinámicamente, en función de su estado y de su entorno, siguiendo un *ciclo de vida* determinado por diversas variables ya comentadas. Pero a su vez, cada uno podrá relacionarse con otros sujetos de su entorno, perteneciendo y formando *grupos* de individuos. Representando los profundos vínculos que se articulan en los colectivos sociales, *sus integrantes se influirán entre sí* teniendo en cuenta las citadas tendencias probabilísticas propias de las ciencias sociales, definidas como fórmulas relacionales que pretenden recoger la múltiple interdependencia de las numerosas variables. Gracias al comportamiento autónomo y flexible de los agentes, a pesar de estar sujetos a las continuas presiones del sistema, se puede observar un comportamiento emergente del conjunto de ellos, cuyas dinámicas de evolución pueden ser estudiadas en sus dimensiones tanto *espacial* (con la extensión de los vínculos grupales) como *temporal* (atendiendo a su demografía).

En esta línea, se puede estudiar, por ejemplo, la influencia religiosa de un grupo especial, la familia, sobre los individuos jóvenes, y cómo ésta va teniendo menos peso sobre ellos a medida que pasan los años. Con el tiempo los agentes, sometidos a mayores influencias y envueltos en otros grupos, podrían relegar a un segundo plano la determinante presión familiar. El conjunto de comportamientos de la misma generación sería expresado por gráficas auto-generadas por el programa simulador, que al ser analizadas e interpretadas permiten, mediante inducción, corroborar (o reformular) hipótesis teóricas. La simulación permite así sustituir al método empírico por excelencia: el experimento.

5 Conclusiones

Con el proceso y herramientas aquí descritos lo que se ha logrado es, esencialmente, eliminar la necesidad de codificar los modelos de simulación social con un lenguaje de programación. En vez de esto, el usuario crea sus modelos de forma gráfica y trabaja con conceptos de un nivel de abstracción mayor. Además, se dispone de un conjunto de herramientas para verificar el cumplimiento de propiedades específicas por los modelos. Estas herramientas se pueden extender si se requiere verificar nuevas propiedades (aunque para ello hace falta la intervención de un ingeniero informático que sepa utilizar la interfaz de programación del IDK para creación de módulos). El usuario sólo tiene que conocer el entorno de simulación y las herramientas que le ofrezca para generar resultados (por ejemplo, diagramas y gráficos de la evolución del sistema).

Otra ventaja de este planteamiento es que se puede hacer replicación de las simulaciones en distintos entornos. Como el modelo se realiza utilizando un lenguaje de modelado gráfico y luego se transforma en código, realizando la transformación para varios entornos de simulación se podrán comparar los resultados posteriormente. En este sentido ya hemos realizado experimentos con RePast y Mason [7], donde hemos estudiado los efectos que pueden tener distintas estrategias de planificación de la simulación. Una vez que se hayan desarrollado varios módulos de generación de código, la simulación de un modelo abstracto cualquiera en diversos entornos será trivial.

El IDK no proporciona módulos de interpretación de los resultados obtenidos con

el simulador. Consideramos que la presentación de resultados directamente en los entornos de simulación es suficiente para el propósito del sociólogo siempre que los elementos del modelo se hayan correspondido adecuadamente con elementos del código del modelo en el simulador (por ejemplo, utilizando los mismos nombres para los agentes en el modelo y en el simulador). De esta manera los diagramas de resultados que proporciona el simulador son fácilmente interpretables por el usuario.

Es posible también hacer una simulación directa sobre el modelo con el IDK, utilizando código generado sobre la plataforma de agentes JADE, pero su ejecución se realiza paso a paso controlada por el usuario. Esto puede resultar tedioso pero es útil para depurar y validar el modelo (comprobar que hace lo que se espera que haga). No nos hemos planteado en el IDK hacer un simulador más potente. Nos ha parecido más interesante reutilizar y facilitar la replicación sobre los entornos existentes, ya bastante avanzados.

Respecto al lenguaje de modelado, la definición de lenguajes más orientados a cada dominio de estudio sería más apropiada que la aplicación directa de INGENIAS. Para ello tenemos previsto definir con equipos de sociólogos cómo serían estos lenguajes y personalizar el editor del IDK para cada uno.

Agradecimientos

Este trabajo ha sido realizado en el contexto del proyecto ”*Métodos y herramientas para modelado de sistemas multiagente”*, subvencionado por el Ministerio de Educación y Ciencia, referencia TIN2005-08501-C03-01.

Referencias

1. Andrés Orizo, F. Y Elzo, J. (Eds): *España entre el localismo y la globalidad. La Encuesta Europea de Valores en su tercera aplicación*, 1981-1999. SM, Madrid, 2000.
2. Arroyo Menéndez, M.: *Cambio cultural y cambio religioso, tendencias y formas de religiosidad en la España de fin de siglo*. Servicio de Publicaciones de la UCM. Madrid, 2004.
3. Bratman, M.E.: *Intentions, Plans and Practical Reason*. Harvard University Press, 1987.
4. Giret, A., Botti, V., and Valero, S.: MAS Methodology for HMS. En: *Holonic and Multi-Agent Systems for Manufacturing, HoloMAS 2005*. Lecture Notes in Artificial Intelligence, 3593. Springer-Verlag (2005) 39—49.
5. OMG: *Meta Object Facility (MOF) Specification. Version 1.4* (2002) formal/02-04-03.
6. Pavón, J., Gómez-Sanz, J.J. & Fuentes, R.: The INGENIAS Methodology and Tools. En: *Agent-Oriented Methodologies.* Idea Group Publishing (2005), 236—276.
7. Sansores, C. y Pavón, J.: Agent-Based Simulation Replication: A Model Driven Architecture Approach. En: *4th Mexican International Conference on Artificial Intelligence (MICAI 2005).* Lecture Notes in Artificial Intelligence, 3789. Springer-Verlag (2005) 244—253.
8. Sansores, C., Pavón, J. y Gómez-Sanz: Visual Modeling for Complex Agent-Based Simulation Systems. En: *Int. Workshop on Multi-Agent-Based Simulation 2005, MABS 2005.* Lecture Notes in Artificial Intelligence, 3891, Springer-Verlag (2006) 174—189.

La aplicación de modelos de consciencia artificial en los sistemas multiagente

Raúl Arrabales Moreno y Araceli Sanchis de Miguel

Departamento de Informática
Universidad Carlos III de Madrid
raul.arrabales@alumnos.uc3m.es, masm@inf.uc3m.es

Resumen. Durante la última década han aparecido algunas implementaciones de modelos científicos de la consciencia basados en sistemas multiagente. El propósito de este artículo es recopilar y describir estos sistemas, determinando hasta que punto estas implementaciones satisfacen los modelos correspondientes, y analizando si proporcionan realmente las supuestas ventajas de usar consciencia artificial en la resolución de problemas. También se analizan en general las funciones de la consciencia y los beneficios que éstas pueden aportar en el rendimiento de los sistemas multiagente.

Palabras clave: Consciencia artificial, sistemas multiagente, atención

1 Introducción a la consciencia artificial

A lo largo de los siglos filósofos, neurocientíficos y psicólogos han desarrollado teorías acerca de la consciencia humana. A pesar de este gran esfuerzo histórico en la búsqueda de una explicación para la consciencia natural, se ha realizado relativamente poco esfuerzo durante las últimas décadas en el campo correspondiente de la inteligencia artificial. Los avances científicos logrados en el estudio de la consciencia durante los 80, que en gran medida aún siguen vigentes, han tenido una influencia modesta en los sistemas artificiales bio-inspirados. A menudo, la consciencia se define basándose en la relación existente en los humanos entre los siguientes procesos mentales: atención, razonamiento, reconocimiento y comportamiento (Kozma, 1997). Es decir, un ser consciente presenta la capacidad de atención hacia una cosa, y puede pensar acerca de ella, qué es, cómo es, por qué es así, etc., con el objetivo de reconocerla. Una vez que el objeto se identifica, el sujeto lo ha reconocido y entonces decide qué quiere hacer con él. Se dice que en los humanos todos estos mecanismos tienen lugar de forma consciente, son contenidos conscientes en nuestra mente. El paradigma de la consciencia artificial se inspira en estos procesos observados en los humanos y otros mamíferos superiores con el objetivo de conseguir sistemas artificiales que presenten capacidades y funcionalidades análogas a las naturales.

Un problema clave en el proceso de modelado computacional de las teorías actuales viene determinado directamente de la propia naturaleza de estas teorías. Algunos de los paradigmas que se aplican para explicar los procesos conscientes, como la mecánica cuántica (Hameroff y Penrose, 1996), los efectos relativísticos (Rakovic,

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 401-412, 2006.

1990), o la sincronización de las activaciones neuronales (Crick y Koch, 1990) son prácticamente imposibles de representar mediante un modelo plausible con agentes, o cualquier otra técnica software. Sin embargo, como explicamos más adelante, existen dimensiones de la consciencia que pueden explicarse con otro tipo de teorías que se prestan mejor a su aplicación y experimentación en sistemas artificiales. Es la dimensión fenomenológica de la consciencia la más incierta en los estudios científicos.

Existen teorías de la consciencia y la atención basadas en parte en aspectos funcionales cognitivos (Baars, 1988; Dennett, 1991; Searle, 1992; Block, 1995; Chalmers, 1997; Sun 2002), las cuales se podrían implementar de forma pragmática por medio de agentes software. Conviene, en cualquier caso, hacer una diferenciación inicial entre dos grandes dimensiones de los procesos conscientes. De esta forma se pueden establecer más claramente los aspectos de la consciencia que se pretenden emular en los sistemas artificiales, sin entrar en el vasto campo de la consciencia humana en toda su extensión e implicaciones. Por supuesto, la tarea de diferenciar entre tipos, clasificaciones y niveles de consciencia no es en absoluto trivial. De hecho, existen numerosas divisiones y clasificaciones (Edelman, 1992; Panksepp, 2005). Se habla de consciencia primaria, autoconsciencia, consciencia afectiva, propiocepción, intersubjetividad, etc. No obstante, existe una clasificación de más alto nivel que es útil para distinguir básicamente entre lo que actualmente podemos aspirar a reproducir en una máquina y lo que no. Se trata de diferenciar entre la consciencia fenoménica (CF) o intransitiva y la consciencia de acceso (CA) o proposicional (Villanueva, 2003). Muchos autores (Block, 1995; Chalmers, 1997; Sugiyama, 2000) sostienen que se puede distinguir entre estos dos tipos de consciencia, o el uso del término consciencia en dos situaciones diferentes. La CF es un tipo de experiencia subjetiva que el sujeto tiene por el hecho de ser consciente, mientras que la CA es de alguna manera la disponibilidad para el uso del razonamiento y la guía de las acciones y el habla. Por lo tanto, se distinguen dos formas de ver la consciencia, por un lado un sujeto es consciente cuando presta atención a un objeto del exterior, conociéndolo y comprendiéndolo mientras éste es foco de su atención; por otro lado, el sujeto puede percibir y sentir su propio interior al tener una experiencia consciente.

Los aspectos de acceso de la consciencia son muy interesantes en cuanto a su posible aplicación en sistemas artificiales. Por otro lado, los aspectos fenoménicos, cuyas características son de una naturaleza poco comprendida hasta el momento, se consideran fuera del ámbito del presente análisis. La consciencia puede ser considerada como una pasarela que proporciona acceso a casi cualquier contenido de la mente (Baars, 1988). Es en definitiva a lo que cada sujeto puede tener acceso sobre si mismo. En un momento dado hay un gran número de procesos neuronales inconscientes ejecutándose en paralelo; sin embargo, sólo ciertos contenidos se muestran a la consciencia en cada momento. Es decir, la atención establece los contenidos de la mente que se perciben conscientemente. Por lo tanto, una de las características principales de los procesos conscientes, en relación a los procesos inconscientes, es que los primeros son mucho más limitados. Los mecanismos conscientes se basan en la memoria a corto plazo y la selección del foco de atención. Estos aspectos son claramente limitados, en el sentido de que no se pueden realizar simultáneamente dos acciones voluntarias (prestar atención a dos cosas a la vez) y la memoria de trabajo que se utiliza no puede manejar más de aproximadamente siete elementos separados al mismo tiempo, por

ejemplo, números de teléfono (Miller, 1956).

Los conceptos vistos anteriormente se pueden emplear para crear modelos computacionales más eficientes, inspirados en el funcionamiento de la consciencia en los mamíferos superiores. En los siguientes apartados pretendemos presentar un breve repaso de las teorías y los modelos cognitivos de la consciencia más importantes, y su estado del arte en cuanto a la implementación y experimentación usando sistemas de agentes.

2 Principales teorías de la consciencia

Existen multitud de teorías sobre el funcionamiento, la evolución, la función y las características de la consciencia. Muchos distinguidos científicos y filósofos, como Nagel, Jackson, McGinn, Damasio, Crick, Dennett, Edelman, etc. han abordado el tema desde diferentes perspectivas. Todas ellas muy interesantes. Sin embargo, en el contexto del presente análisis, y desde un punto de vista pragmático centrado en la aplicación a modelos computacionales, nos centraremos en los trabajos basados en explicar los procesos cognitivos de la consciencia a nivel funcional. En definitiva, como hemos introducido anteriormente, se trata de comprender como se gestiona el acceso al conocimiento y el control de un vasto conjunto de complejos procesos paralelos (inconscientes) desde un único hilo secuencial (consciente). A continuación se describen algunas de las teorías más destacadas, aunque existen muchas más.

La dualidad de la mente

La mayoría de las teorías sobre la consciencia consideran que en la mente existen dos tipos distintos de procesos: conscientes e inconscientes. Desde el punto de vista de los contenidos con los que operan estos procesos, la dualidad se expresa en base a las diferentes formas de representar y procesar el conocimiento. Dependiendo de la naturaleza consciente o inconsciente de los procesos el conocimiento que utilizan puede ser declarativo o procedimental, localizado o distribuido, procesado en serie o en paralelo. Aunque también hay autores que defienden una naturaleza unitaria de lo consciente e inconsciente, como (Dennett, 1991), no existen desarrollos posteriores notorios que hayan desembocado en modelos computacionales. Las hipótesis que consideran la separación entre consciencia e inconsciencia, tienen que plantear los criterios de separación entre ambos dominios, así como el funcionamiento característicos de cada uno de ellos. Por ejemplo, Rosenbloom y Newell (1993) diseñan una arquitectura en la que las tareas se realizan por encadenamiento de diferentes bloques funcionales. Cada bloque es una representación unitaria con un funcionamiento opaco, aunque sus entradas y sus salidas sí son accesibles. La consciencia se produce cuando una tarea se realiza por intervención de múltiples bloques. Si interviene un solo bloque es un proceso inconsciente. Otro criterio establecido por Mathis y Mozer (1996) es que la consciencia se caracteriza por estados temporalmente estables en una red de módulos computacionales especializados. Un criterio básico para la diferenciación entre procesos conscientes e inconscientes es la forma de acceso al conocimiento

(Hadley, 1995; Clark y Toribio, 1992). La representación del conocimiento puede ser implícita o explícita. Los procesos conscientes usan información explícita directamente accesible, mientras que los procesos inconscientes manejan información implícita que no es accesible si no es a través de mecanismos interpretativos.

Sun (2002) argumenta que los procesos cognitivos se estructuran en dos niveles con mecanismos diferentes. Cada nivel codifica un conjunto completo de conocimiento para su procesamiento. Estos dos conjuntos de conocimiento se solapan en gran medida, por lo que los resultados de ambos han de combinarse. Según Sun se produce una sinergia entre el procesamiento implícito (inconsciente) y el procesamiento explícito (consciente). En algunos trabajos sobre la consciencia en robots, como por ejemplo (Kubota et al. 2001) se distingue entre comportamiento automático inconsciente, y comportamiento consciente, que requiere que el individuo tenga cierto grado de autoconsciencia. Denominan autoconsciencia a la consciencia que se tiene de uno mismo y su situación en el mundo. Estos autores establecen dos suposiciones acerca de la autoconsciencia: se activa cuando se produce un cambio relativamente grande en la información sensorial (percepción) y cuando la predicción acerca de la información sensorial es diferente de lo que en realidad se percibe.

La coherencia de la consciencia

Otro tipo de teorías se basan en el concepto de coherencia o coalición. En el caso de (Baars, 1988) una serie de procesadores especializados inconscientes proporcionan información a un espacio de trabajo común, que coordina a los procesadores mediante la selección de patrones coherentes de información (por su valor ilustrativo, analizaremos en más detalle el modelo de Baars a continuación). La noción de coherencia se emplea también a otros niveles, por ejemplo para denotar la activación neuronal sincronizada. En su búsqueda de las correlaciones neuronales de la consciencia, Crick y Koch (1990) llegaron a argumentar que la activación sincronizada a 40 Hz de coaliciones de neuronas era la base física de la consciencia. Aunque más tarde se retractaron al comprobar que estas activaciones entre 35 y 75 Hz en el cortex cerebral no tenían que estar necesariamente relacionadas con los procesos conscientes. También Damasio et al. (1990) hablan de coherencia en otro sentido similar. La reverberación en zonas neuronales de convergencia sensorial integra la información de cada sentido. A su vez, toda la información procedente de cada sentido se integra en una única zona de convergencia multimodal que daría lugar a los contenidos conscientes. De forma similar Schacter plantea la teoría de que múltiples módulos especializados mandan información a un único módulo consciente. En (Schacter et al. 2002) se analizan evidencias de la disociación de los diferentes tipos de conocimiento en el cerebro.

El teatro de la consciencia

Baars (1997) habla de un "teatro", en el que el foco de la consciencia se representa por el punto de luz sobre el escenario, que es dirigido por la atención. El escenario completo se corresponde con la memoria de trabajo, que es el sistema de memoria que almacena los contenidos conscientes. La información obtenida en el punto de luz

se distribuye de forma global a través del teatro a dos clases de procesadores inconscientes: los que forman la audiencia reciben información del foco de luz; mientras, entre bastidores, los sistemas inconscientes contextuales dan forma a los sucesos que ocurren en el punto de luz. La metáfora del foco luminoso es también utilizada por Crick (1994) argumentando, acerca del procesamiento de la información visual, que fuera del punto de luz de la atención visual la información se procesa menos, de forma diferente o ni siquiera se procesa.

No hay que confundir esta metáfora del teatro que usa Baars con otra metáfora denominada "Teatro Cartesiano", que es en esencia opuesta a la defendida por Baars, ya que atribuye la consciencia a un punto concreto del cerebro, la glándula Pineal. Descartes pensaba que en esta glándula se localizaba el alma (Finger, 1995). Las teorías como esta que localizan la consciencia en un punto concreto del cerebro son mayoritariamente rechazadas por la comunidad científica. Si bien es cierto que los neurocientíficos buscan las correlaciones neuronales de la consciencia, no se cree que se localicen en un punto concreto, sino que posiblemente se formen a partir de coaliciones de neuronas (Crick y Koch, 2003).

Volviendo a la metáfora del teatro desarrollada por Baars, es importante resaltar que el "escenario" está compuesto por la memoria de trabajo. Donde los "actores" compiten por aparecer en el foco luminoso de la atención, en el cual aparecen como contenidos completamente conscientes. La selección del foco de atención se realiza en gran medida entre bastidores. Son los procesadores inconscientes los que llevan a cabo esta selección en base al contexto y a conjuntos de creencias (a menudo inconscientes) que determinan los pensamientos conscientes (la actuación en escena). Baars también indica que el foco luminoso de la consciencia es el instrumento que usa el "director" para tomar decisiones en el campo de la memoria de trabajo guiadas por la persecución de metas. Este director de la obra, también trabaja entre bastidores, lo que sugiere que en gran medida no tenemos acceso a las razones por las que hacemos las cosas. Este concepto encaja con el presentado por algunos autores (Rosenthal, 2000; Morin, 2002), que afirman que el *yo* consciente confabula para deducir las razones por las que el sujeto lleva a cabo sus acciones.

Según Baars, al vasto dominio inconsciente de conocimiento y control se puede acceder usando la consciencia. La consciencia se usa para el aprendizaje rápido y el reconocimiento preciso. También activa un gran número de rutinas automáticas que constituyen acciones específicas, proporcionando coordinación y control. Las experiencias conscientes activan contextos inconscientes, que ayudan a interpretar sucesos conscientes futuros. En definitiva, la consciencia proporciona un marco para el acceso (función de búsqueda global) a los vastos contenidos inconscientes de la mente. Parece que las investigaciones realizadas con métodos de diagnóstico por imágenes (resonancia magnética funcional, tomografía por emisión de positrones, etc.) indican que esta hipótesis podría ser cierta (Baars, 2002; Baars et al., 2003); en cualquier caso, se necesitan más análisis neurológicos para confirmar o desmentir con seguridad las suposiciones de Baars.

La dimensión sentimental

Los sentimientos son el balance consciente de nuestra situación (Marina, 2002). Según diferentes teorías psicológicas, los sentimientos son la forma en que los seres humanos son capaces de sintetizar su situación en el mundo dentro del ámbito limitado de la consciencia. El comportamiento está condicionado por el estado emocional del individuo. Como se indica en (Franklin et al. 1998) los sentimientos o emociones proporcionan una valoración de lo bien que se están cumpliendo los objetivos del sistema. Un resumen extraordinariamente simplificado, sin entrar en el complejo mundo de los sentimientos, sería el siguiente: el sujeto sentiría alegría en caso de ver que sus objetivos se van cumpliendo de la forma prevista. En caso contrario, se sentiría frustrado. En los humanos los sentimientos influyen en la conducta, entre otras cosas, en base a un sistema de creencias. Por eso, bajo un estado de frustración unos individuos actúan desistiendo de sus objetivos originales por completo, mientras que otros optarán por intentar diferentes alternativas.

3 Aspectos funcionales de la consciencia

De las teorías sobre la consciencia analizadas hemos tratado de extraer una serie de funciones que creemos son la base de lo que hemos denominado consciencia de acceso (CA). En definitiva se trata de separar la parte funcional de la parte fenomenológica de la consciencia, e identificar en la primera las funciones y características que convierten a la consciencia en una ventaja evolutiva para los seres que la poseen. Hemos identificado las siguientes funciones básicas: (1) atención, (2) balance de situación, (3) búsqueda global, (4) Procesamiento de conocimiento implícito y explícito, (5) contextualización, (6) predicción sensorial, (7) memorización modal y multimodal y (8) autocoordinación. Nuestro planteamiento es que estas funciones deben estar integradas en un sistema de consciencia artificial para que éste presente las ventajas esperadas. En esta lista no hemos incluido otras funciones que se suponen necesarias en un sistema inteligente, pero que no están directamente relacionadas con la consciencia. Como por ejemplo, sistemas de creencias, razonamiento, reconocimiento, percepción, etc. Los detalles sobre la integración de los mecanismos de consciencia artificial con otros paradigmas de la inteligencia artificial están fuera del ámbito del presente análisis. A continuación se describen en más detalle las funciones de la consciencia identificadas:

(1) El mecanismo de atención proporciona al sujeto la capacidad de "prestar atención" a un determinado suceso u objeto, y de esta manera dirigir su aprendizaje y comportamiento.
(2) El balance de situación se refiere a que el sujeto sea capaz de mantener un resumen consciente de su estado. Los sentimientos juegan este papel.
(3) La capacidad de búsqueda global implica acceso a prácticamente todo el conocimiento que posee el sujeto. Esta función es necesaria para el acceso a las rutinas inconscientes de control y las diferentes memorias.

(4) La separación entre procesos conscientes e inconscientes ha de basarse en la diferenciación entre conocimiento explícito e implícito respectivamente. En cada uno de estos dominios debe existir capacidad de aprendizaje. Es decir, aprendizaje implícito inconsciente y aprendizaje explícito consciente. Ambos dominios deben coordinarse a través de mecanismos de control y acceso, como la atención y la búsqueda global.

(5) La contextualización es necesaria para el reclutamiento de procesadores inconscientes adecuados por parte del control consciente. Asimismo, para la resolución de problemas es necesario localizar el conocimiento relacionado con la cuestión que hay que resolver. La memoria asociativa es parte de los mecanismos de contextualización.

(6) La predicción sensorial se refiere a un constante proceso de monitorización y predicción inconsciente de la información obtenida por los sentidos. Cuando lo percibido es distinto de lo esperado, la información correspondiente debe hacerse consciente para poder tratar una situación imprevista.

(7) La memoria multimodal se corresponde con la memoria semántica y de trabajo, donde convergen temporalmente todos los contenidos conscientes. Las memorias modales mantienen indefinidamente contenidos de un tipo específico (por ejemplo, la memoria visual).

(8) La autocoordinación sustituiría en un sistema artificial al libre albedrío. Es decir, se encarga de coordinar las acciones para la consecución de las metas establecidas. Mecanismos como el habla interior y la introspección se incluyen en esta función como elementos de gestión de proyectos (entendiendo por proyecto el conjunto de tareas que se realizan para la consecución de una o varias metas).

Existen diversas sinergias entre las funciones descritas anteriormente, que en su conjunto dan lugar a lo podríamos denominar consciencia artificial. Con respecto a la función 4, hay que remarcar que las novedades requieren mayor participación de la consciencia para su aprendizaje. Es decir, interrelación con la función 1 y 6. Las funciones 5 y 6 representan el flujo de control de arriba abajo y de abajo arriba respectivamente. Por un lado la voluntad consciente invoca procesamientos inconscientes para llevar a cabo sus metas; por otro lado, los procesadores inconscientes que integran la información de los sentidos "llaman la atención" consciente en caso de encontrarse con una situación inesperada o novedosa. Creemos que las propiedades de coherencia o coalición de procesos expresadas por varias teorías se cubren con la función 5, ya que la asociación de procesadores constituye un tipo de coherencia a nivel funcional. La función 7 tiene que dar soporte a la "historia personal del sujeto", que es un aspecto de la consciencia indicado por varias teorías. Este concepto proporciona la unidad necesaria que permite al individuo gestionar su propia experiencia e identidad. El mecanismo de coordinación indicado en la función 8 se relaciona con la función 2 (los sentimientos), ya que la selección de metas y acciones estará condicionada por el estado emocional.

4 Modelos de consciencia implementados con agentes software

Además del caso de las redes de neuronas artificiales, los sistemas multiagente parecen particularmente buenos candidatos para implementar modelos de consciencia porque presentan similitud con el estilo de funcionamiento del cerebro, en el que el trabajo se realiza de forma distribuida por grupos de neuronas especializadas, sin que exista un centro específico de control. Existen diversos sistemas artificiales bio-inspirados en los mecanismos de la consciencia, por ejemplo: *Unified Model of Attention* (Hunt y Lansman, 1986), *ACT* (Anderson, 1993), *ACT-R* (Anderson, 1996), *IDA* (Franklin et al., 1998), *Reflection* (Sugiyama, 2000), *CLARION* (Sun, 1997; Sun, 2002), *CODAM* (Taylor, 2003), *Computational Agent Framework for Consciousness* (Moura y Bonzon, 2004), *SOAR* (Laird et al. 1987; Lehman et al. 2006).

De todos estos modelos, sólo IDA (*Intelligent Distribution Agent*), SOAR y CAFC (*Computacional Agent Framework for Consciousnes*) están implementados mediante sistemas multiagente. Los demás están basados en otro tipo de arquitecturas, como reglas de producción, redes semánticas, redes de neuronas, etc. Como hemos visto, una de las teorías de la consciencia más significativas en el marco de su posible aplicación a los sistemas multiagente es la teoría del Espacio de Trabajo Global (GWT - *Global Workspace Theory*) (Baars, 1988; Baars, 1997). De hecho, tanto IDA como CAFC están basados en esta teoría. Ambos modelos computacionales aplican la hipótesis de Baars acerca de que la consciencia es un suceso global que se produce en partes distribuidas del cerebro, lo cual encaja bien con el concepto de sistema multiagente. Agentes inteligentes independientes se envían mensajes unos a otros a través de un espacio de trabajo común. En este entorno la experiencia consciente emerge de la cooperación y la competición. IDA no es un sistema de propósito general sino que está específicamente diseñado para la optimización en la asignación de tareas a soldados de la marina de los EEUU. Por lo tanto la experimentación está limitada a la interacción que permiten sus interfaces específicos. CACF es potencialmente de propósito general, al igual que SOAR, pero implementa un modelo más limitado de consciencia. Por otro lado, SOAR no implementa directamente un modelo de consciencia, sino que está basado en modelos cognitivos clásicos. Considera los conceptos de meta, estado y operador para manejar el conocimiento.

5 Evaluación de las implementaciones de consciencia artificial

La consciencia en si misma se puede analizar desde muchas perspectivas, y consecuentemente los sistemas de consciencia artificial se pueden evaluar de formas muy diferentes dependiendo de los factores que se consideren relevantes. Básicamente, la ventaja principal que teóricamente se puede obtener por el hecho de aplicar un modelo de consciencia, es el que sistema artificial sea capaz de manejar mejor situaciones nuevas y problemáticas (no esperadas). Uno de los objetivos del presente análisis es dilucidar si la presencia de las funciones de la consciencia mencionadas anteriormente influyen positivamente en el rendimiento del sistema, tal y como se predice en las teorías de la consciencia. Planteamos un método de evaluación basado en dos partes

fundamentales:

- En primer lugar, para cada implementación considerada pretendemos determinar hasta que punto reproduce ésta la correspondiente teoría, analizando las posibles deficiencias. Concretamente, se analizan las funciones de la consciencia que se implementan en cada caso (sean o no consideradas por la teoría que inspira el modelo). Para ello utilizamos la lista de funciones que hemos identificado como clave para un sistema consciente.
- En segundo lugar, se analiza el rendimiento de las implementaciones de los modelos de consciencia. Este análisis debe estar orientado especialmente a determinar la capacidad de los sistemas a enfrentarse con situaciones inesperadas y aprender de las mismas.

La siguiente tabla resume la comparativa de los sistemas multiagente exclusivamente en base a la implementación de las funciones que consideramos clave en un sistema artificial consciente. Queda de manifiesto que ningún sistema abarca todas las funciones.

Tabla 1. Comparación de implementaciones de modelos de consciencia con agentes.

Función Implementada	IDA	CAFC	SOAR
(1) Atención	Sí	Sí	No
(2) Balance de situación	Sí	No	No
(3) Búsqueda global	Sí	Sí	Sí
(4) Procesamiento de conocimiento implícito y explícito	Sí	Sí	No
(5) Contextualización	Sí	Sí	Sí
(6) Predicción sensorial	No	No	No
(7) Memorización modal y multimodal	Sí	Sí	Sí
(8) Autocoordinación	No	No	No

En CAFC se consideran los conceptos de plan y condición, pero no existe un módulo de coordinación como tal que dirija la construcción de planes. Particularmente, como no existe la función de balance de situación, no es posible dirigir los planes de acuerdo al progreso que se está realizando. Hay que evaluar frente a diferentes problemas, para comprobar efectivamente la robustez respecto a problemas y situaciones nuevas así como la flexibilidad del sistema. Esto no se analiza enfrentándose siempre al mismo dominio de problemas. Es necesario realizar una correlación entre funciones, grupos de funciones, y su impacto en el rendimiento.

5 Conclusiones

Las hipótesis barajadas en el presente artículo se basan principalmente en metáforas que simplemente ayudan a comprender de forma holística el funcionamiento de la mente humana. Si bien es cierto que una simple metáfora está muy lejos de constituir un cuerpo establecido de conocimiento científico, puede servir como herramienta para dirigir las investigaciones en diversos sentidos, que afirmen o desmienten las hipótesis planteadas. Desde el punto de vista de la implementación de modelos computacionales, el uso de estos esquemas simplificados de la consciencia tiene dos ventajas claras: la facilidad de implementación y comprensión del modelo, y por ende la posibilidad de experimentación con sistemas artificiales fácilmente observables, parametrizables y relativamente asequibles. Por supuesto en ningún caso la experimentación con sistemas artificiales puede en modo alguno sustituir a la experimentación con los verdaderos poseedores de consciencia natural. Sin embargo, como sabemos, la inteligencia artificial ha ofrecido retroalimentación útil durante los últimos 50 años a la psicología y viceversa, completando y mejorando los modelos de la mente en base a los resultados obtenidos en ambos dominios. Pensamos que en el campo de la consciencia, que no es más que una parte que ha de integrarse con las teorías existentes de la mente, ha de suceder lo mismo. Un modelo de consciencia no es suficiente, ya que la consciencia es un aspecto fundamental, pero no el único. Se requiere un modelo completo de la mente. En cualquier caso, conocer como funciona la consciencia y sus funciones asociadas es un buen comienzo para encajar el resto de piezas del puzzle. Aunque las teorías analizadas cubren un amplio rango de funcionalidades, uno de los conceptos que echamos en falta es el concepto de proyecto. Creemos que esta noción, entendida como la asociación de metas orientadas a conseguir un objetivo final, debe ser considerada como parte de la funcionalidad de autocoordinación.

La consciencia puede ser la forma que el sistema nervioso ha desarrollado evolutivamente para lidiar con sucesos novedosos e inesperados en el mundo (Franklin, 2005). Esta concepción supera los antiguos sistemas situacionales, como por ejemplo la arquitectura de subsunción (Brooks, 1990), en los que sin necesidad de controlar estados o conocimientos internos, un agente autónomo puede desenvolverse de manera exclusivamente reactiva. Un ejemplo de las carencias de este tipo de sistemas es (Arrabales et al., 2002), donde la necesidad de un mecanismo de atención se hace patente. El método de evaluación planteado, aunque se trata de una primera aproximación, proporciona medidas heurísticas acerca de que funciones de la consciencia son clave en el aumento de rendimiento (entendiendo por rendimiento, la mejor capacidad adaptativa del sistema). Ha de tenerse en cuenta la imposibilidad de realizar este tipo de evaluaciones con sistemas naturales. Por razones obvias no se pueden añadir y quitar funciones mentales a un sujeto natural (ya sea un ser humano u otro animal). Experimentar con sistemas diferentes también es un problema. Habría que probar con el mismo sistema añadiendo y quitando componentes (funciones). Esto se plantea como un trabajo futuro: un sistema modular (tipo banco de pruebas) con el que se pueda evaluar mejor la aportación de todas y cada una de las funciones de la consciencia (y la correspondiente integración y sinergias entre diferentes funciones).

Referencias

ANDERSON, J.R. (1993). *Rules of Mind*. Hillsdale. Lawrence Erlbaum.

ANDERSON, J.R. (1996). *The Architecture of Cognition*. Lawrence Erlbaum.

ARRABALES, R. FLANAGAN, C. y TOAL, D. (2002). *An Adaptive Video Event Mining System for an Autonomous Underwater Vehicle*. Intelligent Engineering Systems through Artificial Neural Networks, Vol. 12, pp. 585-591. ASME Press.

BAARS, B.J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.

BAARS, B.J. (1997). *In the Theater of Consciousness. Global Workspace Theory, A Rigorous Scientific Theory of Consciousness*. Journal of Consciousness Studies, 4, No. 4, 1997, pp. 292-309.

BAARS, B.J. (2002). *The conscious access hypothesis: origins and recent evidence*. Trends in Cognitive Sciences, Vol. 6 No. 1 pp. 47-52.

BAARS, B.J. RAMSOY, T.Z. y LAUREYS, S. (2003). *Brain, conscious experience and the observing self*. Trends in Neurosciences. Vol. 26, No. 12, pp. 671-675.

BLOCK. N. (1995). *On a Confusion about a Function of Consciousness*. Behavioral and Brain Sciences 18, 227-87.

BROOKS, R.A. (1990). *Elephants Don't Play Chess*. Designing Autonomous Agents. MIT Press.

CHALMERS. D. (1997). *Availability: The Cognitive Basis of Experience*. The Nature of Consciousness, Edited by Block. N., Flanagan. O. and Guzeldere. G., MIT Press.

CLARK, A. y TORIBIO, J. (1994). *Doing without Representing?* Synthese, vol. 101, No. 3. Springer Science.

CRICK, F. (1994). *Astonishing Hypothesis: The Scientific Search for the Soul*. Scribner Book Company.

CRICK, F. y KOCH, C. (1990). *Toward a neurobiological theory of consciousness*. Seminars in the Neurosciences 2:263-275.

CRICK, F. y KOCH, C. (2003). *A framework for consciousness*. Nature Neuroscience, 6:119-126.

DAMASIO, A.R. DAMASIO, H. TRANEL, D. y BRANDT, J.P. (1990). *Neural Regionalization of knowledge access: preliminary evidence*. 55, pp. 1039-1047. Cold Spring Harbor Symposia on Quantitative Biology.

DENNETT, D.C. (1991). *Consciousness Explained*. Penguin.

EDELMAN, G.M. (1992). *Bright Air, Brilliant Fire. On the Matter of the Mind*. Basic Books.

FINGER, S. (1995). *Descartes and the pineal gland in animals: a frequent misinterpretation*. Journal of the history of the neurosciences. Sep-Dec; 4(3-4): 166-82.

FRANKLIN, S. (2005). *Evolutionary pressures and a stable world for animals and robots: A commentary on Merker*. Consciousness and Cognition 14, pp. 115-118.

FRANKLIN, S. KELEMEN, A. y McCAULEY, L. (1998). *IDA: A Cognitive Agent Architecture*. IEEE Conference on Systems, Man and Cybernetics. IEEE Press.

HADLEY. R.F. (1995). *The "explicit-implicit" distinction*. Minds and Machines, Vol. 5, No. 2. Springer Science.

HAMEROFF, S.R. y PENROSE, R. (1996). *Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness*. Toward a Science of Consciousness - Contributions from the Tucson Conference, MIT Press, Cambridge, MA.

HUNT, E. y LANSMAN, M. (1986). *Unified Model of Attention and Problem Solving*. Psychological Review, 4. pp. 446-461.

KOZMA, R. (1997). *On the conscious and subconscious components of knowledge representation in neural networks*. International Conference on Neural Networks. Vol. 4, pp. 2519-2523.

LAIRD, J.E. NEWELL, A. y ROSENBLOOM, P.S. (1987). *SOAR: an architecture for general*

intelligence. Artificial Intelligence, vol. 33, pp. 1-64. Elsevier Science.

LEHMAN, J.F. LAIRD, J. ROSENBLOOM, P. (2006). *A Gentle Introduction to Soar, an Architecture for Human Cognition: 2006 Update*. Actualización del original Sternberg y Scarborough (1996).

MARINA, J.A. (2002). *El laberinto sentimental*. Editorial Anagrama.

MATHIS, D.W. y MOZER, M.C. (1996). *Conscious and Unconscious Perception: A Computational Theory*. Proceedings of the Eighteenth Conference of the Cognitive Science Society, pp. 324-328. Erlbaum.

MILLER, G.A. (1956). *The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information*. The Psychological Review, vol. 63, pp. 81-97.

MORIN, A. (2002). *Do you "self-reflect" or "self-ruminate"?* Science and Consciousness Review. Dec. No. 1.

MOURA, I. y BONZON, P. (2004). *A Computational Framework for Implementing Baars' Global Workspace Theory of Consciousness*. Brain Inspired Cognitive Systems.

PANKSEPP, J. (2005). *Affective Consciousness: Core emotional feelings in animals and humans*. Consciousness and Cognition 14 30-80.

RAKOVIC, D. (1990). *Neural Networks Vs. Brain Waves: Prospects for Cognitive Theory of Consciousness*. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 12, No. 3.

ROSENBLOOM, P.S. y NEWELL, A. (1993). *The chunking of goal hierarchies: a generalized model of practice*. The SOAR papers, vol. 1. MIT Press Series in Research In Integrated Intelligence.

ROSENTHAL, D.M. (2000). *Consciousness, Content, and Metacognitive Judgements.* Consciousness and Cognition 9, pp. 203-214

SCHACTER, D.L. REIMAN, E. UECKER, A. ROISTER, M.R. YUN, L.S. y COOPER, L.A. (2002). *Brain regions associated with retrieval of structurally coherent visual information*. Nature, 376, pp. 587-590.

SEARLE, J. (1992). *The rediscovery of the Mind.* The MIT Press.

SUGIYAMA, S. (2000). *Reflected Method for Having Consciousness*. IEEE International Conference on Systems, Man, and Cybernetics, 2000. Volume 5, 8-11 Oct. 2000 Page(s):3141 - 3146 vol.5.

SUN, R. (1997). *Learning, Action and Consciousness: A Hybrid Approach Toward Modelling Consciousness*. Neural Networks, Vol. 10, No. 7, pp. 1317-1331. Elsevier Science Ltd. 0893-6080/97.

SUN, R. (1999). *Computational Models of Consciousness: An Evaluation*. Journal of Intelligent Systems, 9. pp. 507-562.

SUN, R. (2002). *Duality of the mind. A bottom up approach toward cognition*. Lawrence Erlbaum Associates Publishers.

TAYLOR, J. (2003). *The CODAM model of Attention and Consciousness.* Proceedings of the IEEE International Joint Conference on Neural Networks, 2003, July 20-24.

TAYLOR, J.G. (1994). *The relational mind*. From Perception to Action Conference. IEEE.

VILLANUEVA, E. (2003). *¿Qué son las propiedades psicológicas? Metafísica de la psicología*. Instituto de Investigaciones Jurídicas. Universidad Nacional Autónoma de México.

Una arquitectura multi-agente con control difuso colaborativo para un robot móvil

Bianca Innocenti, Beatriz Lopez y Joaquim Salvi

Instituto de Informática y Aplicaciones
Universidad de Girona, España
{bianca,blopez,qsalvi}@eia.udg.es

Resumen Uno de los desafíos actuales en el desarrollo de sistemas de control de robots, es que sean capaces de exhibir respuestas inteligentes y adecuadas a entornos cambiantes. Resulta tecnológicamente difícil y potencialmente peligroso construir sistemas complejos controlados exclusivamente de forma centralizada. Una manera de abordar la descentralización es mediante el control colaborativo, ya que permite desarrollar un comportamiento complejo a partir de diversos controladores que se combinan para conseguir el resultado deseado. Por otra parte, un robot requiere de capacidades cognitivas de más alto nivel. En este caso, las arquitecturas multiagentes proporcionan el modo adecuado para definirlas. Este artículo pretende combinar las arquitecturas multiagentes con el control colaborativo, de forma que éste se desarrolla dentro del agente. Se presenta la arquitectura multiagente y se hace énfasis en la integración del control colaborativo para modelar el comportamiento de un agente. Los experimentos se han realizado en un robot móvil Pioneer.

1. Introducción

Uno de los desafíos actuales en el desarrollo de sistemas de control para robots, es que sean capaces de exhibir respuestas inteligentes y adecuadas a las circunstancias cambiantes del entorno. Métodos de aprendizaje y de adaptación, así como técnicas de toma de decisiones, ayudan en la consecución de estos objetivos. Sin embargo, resulta tecnológicamente difícil y potencialmente peligroso construir sistemas complejos que sean controlados exclusivamente de forma centralizada [16].

Uno de los trabajos pioneros en considerar arquitecturas compuestas de diferentes comportamientos distribuidos, independientes y asíncronos que eran coordinados por un árbitro central ha sido DAMN [22]. El comportamiento global de la arquitectura de Rosenblatt se evaluó como racional, coherente, orientado a objetivos a la vez que mantenía una respuesta en tiempo real, del sistema a su entorno físico inmediato. Otras arquitecturas se han desarrollado desde entonces, como por ejemplo O2CA2 [19], en la que se utiliza un esquema de votación para coordinar los diferentes comportamientos. Las ventajas de este tipo de arquitecturas con comportamientos independientes son que facilitan el desarrollo

A. Fernández-Caballero, M.G. Manzano, E. Alonso & S. Miguel (Eds.): CMPI-2006, pp. 413-424, 2006.

y que permiten la creación evolutiva de sistemas robustos con mayores capacidades [22]. En trabajos más recientes [1], se analizan diferentes arquitecturas y se destaca la organización jerárquica de los comportamientos como una de las organizaciones más ventajosas. En este tipo de arquitecturas, cada comportamiento ha sido modelizado como un módulo con habilidades de comunicación.

Trabajos recientes en Sistemas Multi-Agente, han animado a los investigadores a ir un paso más allá en el diseño de architecturas de control, de manera que los módulos han sido reemplazados por agentes. Los agentes son programas autónomos que pueden interaccionar entre ellos y adaptarse al entorno [15]. Un agente es un sistema auto-contenido, independiente, situado y que tiene capacidades reflectivas y reflexivas (*selfawareness*). Los agentes proporcionan más flexibilidad al desarrollo de arquitecturas para robots. En relación a las restricciones de comunicación que puedan surgir como consecuencia de la interacción de los agentes para su coordinación, trabajos recientes en sistemas multi-agente han demostrado la capacidad de respuesta en tiempo real de estos sistemas a su entorno [25].

Sin embargo, los comportamientos asociados a cada agente, no son sencillos. Por ejemplo, un comportamiento "ir a", en una trayectoria libre de obstáculos, tiene que considerar si el punto de destino está cerca o lejos de la posición actual del robot. Dependiendo de esta información, el robot debería desplazarse rápida o lentamente. Por lo tanto, el comportamiento asociado al agente que forma parte de la arquitectura de un robot, puede, a su vez, ser complejo. Una aproximación al diseño de los agentes es aprovechar los trabajos realizados en control colaborativo para desarrollar un comportamiento complejo a nivel de agente. Este control colaborativo consiste en la combinación de diferentes controladores para obtener el comportamiento deseado.

Integrando ambas líneas de investigación, la aproximación multi-agente y control colaborativo, se obtiene como resultado una arquitectura multi-agente de controles colaborativos, a la que llamamos aproximación de control colaborativo múltiple en un robot. El control colaborativo se aplica al diseño y al desarrollo de un único comportamiento, mientras que toda la arquitectura del robot se basa en un sistema multi-agente en el cual, cada agente representa un comportamiento. En este artículo se describe brevemente la arquitectura multi-agente propuesta y se proporciona una explicación detallada de la implementación de un agente con control colaborativo basado en la lógica difusa [13]. Detalles sobre la aproximación MAS se pueden encontrar en [11]. El objetivo de este artículo, por lo tanto, es centrarnos en el control colaborativo que se realiza en un agente de la arquitectura multiagente.

El artículo se organiza como sigue. En la Sección 2, se presenta el trabajo relacionado. Una breve descripción de la arquitectura multi-agente se proporciona en la Sección 3, que incluye el comportamiento *goto* diseñado por medio del control colaborativo. A continuación, en la Sección 4 se detalla la propuesta de control colaborativo. En la Sección 5 se muestran los resultados obtenidos y finalmente, se presentan las conclusiones y el trabajo futuro en la Sección 6.

2. Trabajo relacionado

Como se estableció anteriormente, el control colaborativo múltiple en un único robot concierne al control colaborativo y a los sistemas multi-agentes. El control colaborativo tiene un significado muy general y por lo tanto, cada vez que se definen varios algoritmos de control en una tarea compleja se está introduciendo una idea implícita de colaboración en el control. Por lo tanto, cualquier desarrollo de sistemas complejos con arquitecturas multiagentes puede considerarse como una aproximación colaborativa. Sin embargo la mayoría de las arquitecturas construidas como sistemas multi-agente en robótica no son para controlar un único robot sino que hacen referencia a sistemas multi-robots ([14],[24],[5],[21]). Las excepciones ([17],[2],[9],[20]) presentan diferentes aproximaciones. Por ejemplo, la arquitectura multi-agente (MAS) propuesta por [17] para el nivel reactivo, tiene dos clases de agentes: los elementales que tienen las habilidades básicas y los de alto nivel que son responsables de integrar y coordinar varios agentes elementales. En [2], hay un agente específico que, basado en un proceso de subastas, determina la acción que debe ejecutarse. En [9] la definición de una arquitectura de sistema de control MAS se basa en la teoría de la organización y en la de las alianzas estratégicas. En [20] se propone un sistema multi-agente que mejora la navegación de robots autónomos en entornos semiestructurados desconocidos. La novedad de este trabajo es el uso de técnicas de razonamiento basado en casos (CBR) para tratar las situaciones problemáticas, como por ejemplo, caminos sin salida o disposición de los obstáculos que el robot no pueda esquivar.

Por otra parte, el control colaborativo se aplica a tres tipos de sistemas: la fusión de sensores, dónde el conocimiento se obtiene procesando la información proporcionada por múltiples fuentes; el control de procesos múltiples, en el que varios controladores comparten el control del sistema; y múltiples operadores humanos, en el cual varios operadores comparten y negocian el control. Nuestra investigación se centra en la rama del control de procesos múltiples.

En [10] varias fuentes controlan cada rueda. Los controladores se modelan como autómatas finitos cuyas entradas son la posición del robot. El incremento deseado en el movimiento, se basa en la media del voto conjunto de las salidas de los autómatas. En [7], un conocimiento de alto nivel se usa para seleccionar el controlador adecuado para realizar la acción deseada. La selección del controlador se hace escogiendo de una base de casos de controladores, el que tiene especificaciones similares a las enviadas por el subsistema de alto nivel.

Otras aproximaciones, como la de [8], muestran como el control colaborativo (llamado control concurrente) presenta algunas propiedades deseables para el control de robots móviles como tolerancia a fallos, distribución y escalabilidad. Esta aproximación se basa en la intercalación de órdenes provenientes de diferentes controladores en los motores. El movimiento resultante se consigue como una superposición de diferentes señales de control.

De acuerdo con [23], nuestra hipótesis es que, en lugar de superponer señales, se puede usar un procedimiento de toma de decisiones de alto nivel para coordinar los diferentes controladores. Particularmente, proponemos el uso de la lógica difusa para modelar las acciones de control provenientes de controladores he-

terogéneos y para decidir, de acuerdo con el modelo dinámico del robot, que combinación de acciones de control debe ejecutarse en cada momento. Nuestra principal diferencia con el trabajo de Saffioti es que no estamos implementando un controlador difuso por medio de reglas difusas, sino que usamos los conjuntos difusos para modelar la relevancia de los diferentes controladores en las medidas de agregación extendiendo el trabajo de [8].

Además, el resultado del controlador colaborativo es la salida de un único comportamiento, en lugar de ser la orden directa de los actuadores. La salida de cada comportamiento (agente), se coordina en la arquitectura multi-agente para decidir las acciones posteriores del robot.

3. La arquitectura de control colaborativo múltiple

La arquitectura MAS propuesta se muestra en la Fig. 1. Los agentes pueden agruparse en agentes de percepción, de comportamientos, deliberativos y actuadores. Los agentes de percepción obtienen la información del entorno y de las condiciones internas del robot; los de comportamiento realizan acciones específicas, como evitar obstáculos; los deliberativos implementan tareas complejas como la planificación; y los agentes actuadores se encargan de controlar la velocidad lineal y angular del robot, interaccionando directamente con los motores.

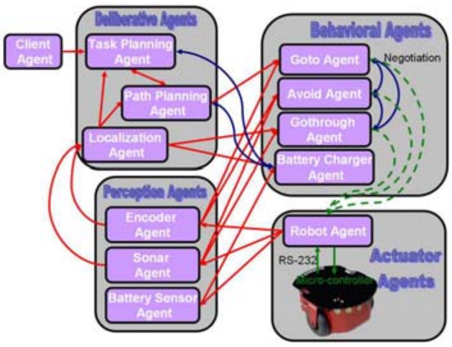


Figura 1. La arquitectura de control colaborativo múltiple así como el flujo de información entre los diferentes agentes. Las líneas de puntos indican que un único agente se comunica con el agente *robot*.

Respecto al agente *goto*, que es el objeto del control colaborativo presentado en este artículo, es un agente de comportamiento responsable de conducir al robot hasta la posición meta (rápida o lentamente) en función de la información que recibe del agente *encoder*.

La coordinación entre los agentes se necesita cuando hay varios que quieren usar el mismo recurso al mismo tiempo. En la arquitectura presentada, estos conflictos pueden aparecer entre los agentes *avoid*, *goto* y *goThrough* cuando intentan enviar acciones contradictorias al agente *robot* (actuador). El agente

avoid es responsable de evitar los obstáculos mientras que el agente *goThrough* conduce el robot a través de lugares estrechos como las puertas.

Una solución a este problema es definir un agente coordinador central, que sabiendo cuales son los agentes en conflicto, impone una decisión. Sin embargo, creemos que este mecanismo de coordinación centralizado puede ser un problema cuando haya muchos agentes en la arquitectura. En cambio, pensamos que si los conflictos son locales, una aproximación distribuida de coordinación puede ser más apropiadas. Concretamente, se propone un mecanismo de coordinación *peer-to-peer* entre los agentes involucrados en el conflicto, basados en el cálculo del valor de la *utilidad*. Para más detalles, ver la explicación presentada en [11].

Por motivos de longitud en este artículo nos centramos en el detalle del agente *goto* diseñado como un control colaborativo. Como agente, es capaz de comunicarse con los otros agentes de la arquitectura para decidir las acciones del robot. En su toma de decisiones interna utiliza diversos controladores.

4. Control Colaborativo

En esta sección se presenta el método de control colaborativo, basado en lógica difusa para hacer la combinación de múltiples controladores, que implementan el comportamiento del agente *goto*. En lugar de desarrollar un único controlador muy elaborado, pretendemos diseñar varios controladores sencillos orientados a tratar diferentes aspectos del control por separado y unir sus acciones para obtener un comportamiento final complejo.

El punto de partida es la ecuación propuesta por [8], que se extiende agregando los pesos correspondientes a la relevancia de cada controlador, de acuerdo con el contexto actual. Así, el cálculo de la velocidad deseada se define como un promedio ponderado de las órdenes proporcionadas por los controladores, siguiendo la fórmula siguiente:

$$\Omega = \frac{\sum_{t=1}^{n} \eta_t \cdot w_t}{\sum_{t=1}^{n} w_t} \qquad (1)$$

donde η_t es la velocidad deseada de las ruedas para cada instante de tiempo, Ω la velocidad real de la rueda, n el número de controladores y w_t los pesos de cada controlador. Estos pesos satisfacen la condición de $\sum_{t=1}^{n} w_t = 1$.

Al usar los pesos, es posible darle más o menos importancia a los diferentes controladores. Para implementar este ajuste de velocidades expresado por la ecuación 1, se propone un sistema difuso funcional de Sugeno [6] (ver Fig. 2).

La entrada del sistema la conforman $n+1$ variables, una por cada controlador, y la distancia d al objetivo. Esta última se define en función de la distancia que ya ha recorrido el robot d_{rec} y la distancia desde la posición inicial del robot hasta el punto destino d_{max}, esto es $d = (d_{max} - d_{rec})/d_{max}$. Se han definido dos posibles valores para d: cerca y lejos. Los términos lejos y cerca se modelan mediante conjuntos difusos (fuzzy sets). El conjunto difuso *cerca* se puede definir como:

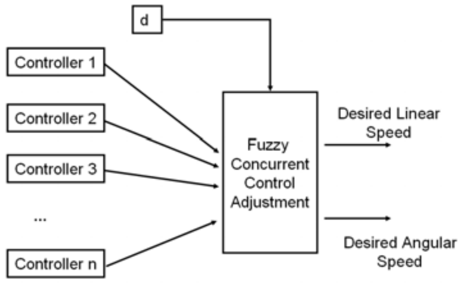


Figura 2. Diagrama de bloques de la aproximación propuesta.

$$\mu(d) = \begin{Bmatrix} 1 & d \leq min \\ \frac{(-d+max)}{(max-min)} & min < d < max \\ 0 & d \geq max \end{Bmatrix} \tag{2}$$

donde min y max son parámetros que se han sintonizado empíricamente.

De acuerdo cone esta definición, dependiendo del movimiento del robot, la distancia al punto destino puede variar en forma no lineal en el tiempo, provocando que los conjuntos fuzzy sean también no lineales. La Fig. 3 muestra diferentes resultados para la función descrita por (2). En la Fig. 3-a) se representa un caso lineal, mientras que a las Fig. 3-b) y 3-c) se revela la no-linealidad de la ecuación, que depende además de los parámetros min y max.

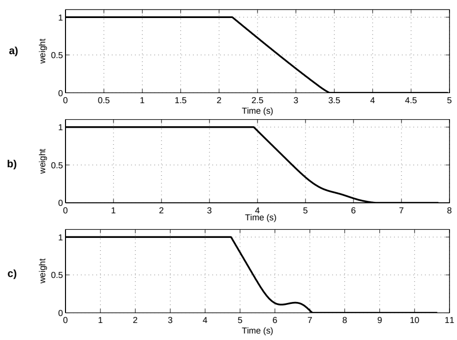


Figura 3. Pesos *fuzzy*.

Por otra parte, el conjunto difuso lejos se define como:

$$\mu_{lejos}(d) = 1 - \mu_{cerca}(d) \tag{3}$$

Como variables de salida tendremos la velocidad lineal y la velocidad angular finales. En el caso de combinar dos controladores, se definen dos valores posibles para dichas variables de salida: *lenta* y *rápida*.

En este sistema de Sugeno con $n+1$ variables, los valores de las variables de salida se determinan como combinación lineal de las entradas. Es decir, para cada valor x de la variable de salida s se define un valor:

$$u_x(s) = a_{n+1} \cdot v_{n+1} + ... + a_1 \cdot v_1 + a_0 = (a_{n+1}, ..., a_1, a_0) \tag{4}$$

donde a_i es el coeficiente de la variable v_i y a_0 una constante.

En nuestro caso, considerando dos controladores se han definido los vectores de salida siguientes: $u_{lenta}(v) = (0, 0, 1, 0)$ y $u_{rapida}(v) = (0, 1, 0, 0)$, donde la variable $n+1$ es la distancia d al objetivo, v_n es el controlador rápido y v_{n-1} es el controlador preciso. La constante a_0 se ha escogido 0.

Las reglas que describen el sistema difuso son:

Si **d** es cerca entonces **v** es lenta
Si **d** es lejos entonces **v** es rápida.

La función elegida para la propagación de evidencias (de premisas a conclusiones) es el producto. Como consecuencia de ello, si se activa la primera regla se obtiene como resultado $\mu_{cerca}(d) \cdot PID_{preciso}$ y si se activa la segunda regla $\mu_{lejos}(d) \cdot PID_{rapido}$, siendo su combinación el resultado que se expresa en la ecuación (1) (al utilizar la función media como el método de defuzzyficación).

5. Resultados

El principal objetivo es obtener una arquitectura de control del robot que proporcione una conducta coherente, racional y dirigida a cumplir el objetivo marcado, preservando a la vez, la respuesta en tiempo real al entorno físico inmediato [22]. En este sentido, los sistemas multi-agentes proporcionan la arquitectura básica, mientras el control colaborativo hace al agente *goto*, encargado de conducir al robot a la posición deseada con la orientación requerida, rápido y preciso.

Para el agente *goto*, se ha decidido implementar dos controladores diferentes y mezclar los dos vectores de control resultantes mediante factores de pesos fuzzy. Un controlador es muy rápido pero no llega exactamente a la consigna, mientras el otro es lento pero logra llegar exactamente a la consigna fijada. Debido a la naturaleza no lineal del robot, es posible que haya algunas consignas a las que no sea posible llegar exactamente con los controladores diseñados, ya que son PIDs comunes. El lazo de control del control colaborativo propuesto se muestra en la figura Fig. 4. El bloque del control concurrente difuso se encarga de la mezcla de las velocidades deseadas de los controladores para conmutar progresivamente de un controlador al otro.

En lo que resta de esta sección, se explican los diferentes controladores desarrollados, los resultados obtenidos utilizando cada controlador por separado y la salida del control colaborativo. Todas las pruebas se han realizado en trayectorias libres de obstáculos y que no ocurren en pasos estrechos. Los resultados, por tanto, reflejan la precisión del comportamiento del robot. El tiempo de respuesta es satisfactorio. Otras evaluaciones que conciernen el coste de coordinación en la arquitectura multiagente están analizados en [11].

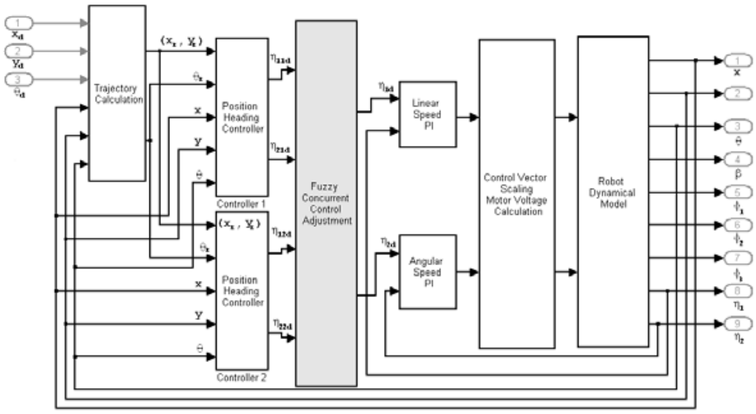


Figura 4. Diagrama de bloques del control colaborativo.

5.1. Controladores de velocidad

Con el objetivo de obtener la velocidad lineal y angular deseadas, se han desarrollado dos controladores PI independientes, uno para la velocidad angular y otro para la lineal. Un aspecto particular de estos controladores es que, en lugar de utilizar la aproximación T (T-approximation, truncado del vector de control) utilizan la aproximación S (S-approximation, escalado) para obtener vectores de control $\mathbf{u(t)}$ físicamente posibles que no violen las restricciones (más detalles en [18]). Ya que el vector resultante es proporcional al no realizable, los dos tienen la misma dirección y por lo tanto no hay error respecto a la orientación deseada. Este proceso se realiza en el bloque *Control Vector Scaling and Motor Voltage Calculation*.

5.2. Controlador de posición rápido

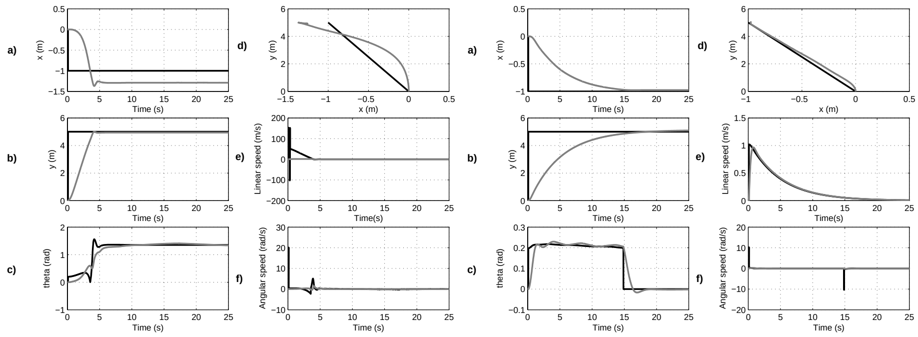
La idea es obtener un controlador de posición que sea muy rápido sin importar si llega exactamente a la consigna deseada o no. La Fig. 5.1 representa la respuesta del sistema controlado para una posición y orientación iniciales de $(x_0, y_0, \theta_0) = (0, 0, 0)$ y una consigna deseada de $(x_f, y_f, \theta_f) = (-1, 5, 0)$. Las Fig. 5.1-a) y 5.1-b) muestran la evolución de las coordenadas x e y con el tiempo, la Fig. 5.1-c) la orientación del robot en función del tiempo, las Fig. 5.1-e) y f) las velocidades lineal y angular respectivamente y finalmente la trayectoria del robot ($y = x(t)$) se representa en la Fig. 5.1-d). Tal como puede verse en la Fig. 5.1-d) el robot no llega exactamente a la posición deseada, incluso ni la orientación es la deseada, aunque lo hace rápido: se para en 5 segundos.

5.3. Controlador de posición preciso

El objetivo ahora es obtener el controlador de posición que llegue a la posición y orientación deseadas, a pesar del tiempo empleado en alcanzar la consigna.

La Fig. 5.2 presenta la respuesta para las mismas posición y orientación iniciales y finales del experimento anterior. Las Fig. 5.2-a) y Fig. 5.2-b) muestran la evolución de las coordenadas x e y con el tiempo, la Fig. 5.2-c) la evolución de la orientación, las Fig. 5.2-e) y Fig. 5.2-f) las velocidades lineales y angulares y la Fig. 5.2-d) la trayectoria del robot ($y = x(t)$).

Como puede observarse, el robot tarda más de 10 segundos en llegar a la consigna. Este tiempo es el doble del empleado por el controlador rápido. No obstante, el controlador rápido presenta un error en la posición final, mientras que el controlador preciso consigue llegar a la consigna con error nulo.



1- PID rápido 2- PID preciso

Figura 5. Respuesta de los controladores PID.

5.4. Control Concurrente Fuzzy

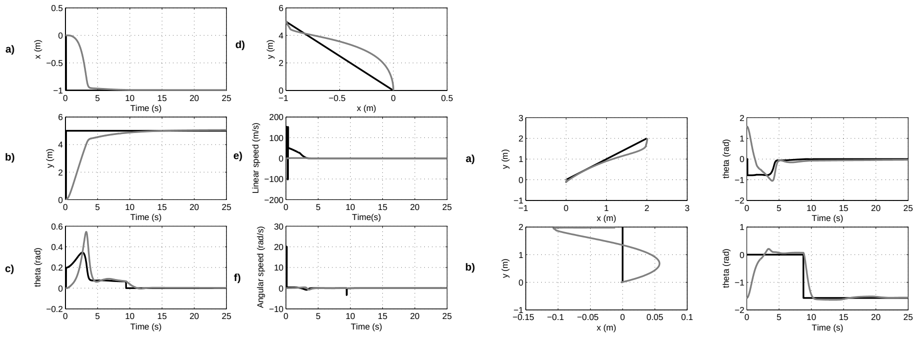
La Fig. 6.1 muestra la respuesta del sistema completo usando control concurrente (ver (1)), para las mismas posiciones inicial y final y orientación que en los experimentos previos. Como puede verse, el robot llega más rápido a la consigna deseada que el controlador lento y con más precisión que el controlador rápido

Se han llevado a cabo otros experimentos que han dado resultados similares. Como ejemplo, la Fig. 6.2 muestra la respuesta del control fuzzy concurrente para diferentes posiciones y orientaciones iniciales, así como finales.

Para la Fig. 6.2-a) el estado inicial es $(x_0, y_0, \theta_0) = (0, 0, \frac{\pi}{2})$ y el estado deseado es $(x_f, y_f, \theta_f) = (2, 2, 0)$; y para la Fig. 6.2-b), $(x_0, y_0, \theta_0) = (0, 0, \frac{-\pi}{2})$ y $(x_f, y_f, \theta_f) = (0, 2, \frac{-\pi}{2})$.

Una característica interesante de este controlador concurrente es que funciona mejor que los controladores por separado, especialmente para aquellos estados producidos por la naturaleza no lineal del modelo del robot.

Por ejemplo, considerando el estado inicial como $(x_0, y_0, \theta_0) = (0, 0, \frac{-3\cdot\pi}{8})$ y como estado final $(x_f, y_f, \theta_f) = (-2, 3, \frac{-\pi}{2})$, puede sólo conseguirse con el controlador concurrente, tal y como se muestra en la Fig.7. Los resultados que se muestran en la Fig. 7-b) para el controlador más rápido son de esperar, ya que



1- Para la posición ejemplo 2- Para diversos puntos destino

Figura 6. Respuesta del control fuzzy concurrente.

fue diseñado para ser rápido pero no preciso. La Fig. 7-a) muestra la respuesta del controlador más lento; aunque intenta ser preciso, no puede llegar a la consigna deseada debido a las no linealidades. Esto se puede resolver mejorando el diseño del controlador (por ejemplo un controlador no lineal) o usando un controlador concurrente, tal y como se muestra en la Fig. 7-c).

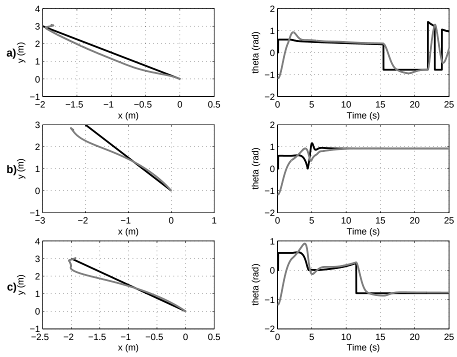


Figura 7. Comparación de la respuesta de los controladores; a) el controlador preciso, b) el controlador rápido y c) el controlador colaborativo.

Finalmente cabe destacar que la experimentación con el robot tiene que extenderse en un futuro para trayectorias planificadas. Los resultados mostrados en esta sección se han realizado para el comportamiento del agente *goto*, que mueve al robot de un punto inicial a un punto final. Con la introducción del planificador de trayectorias el agente *goto* recibirá las coordenadas objetivos de los diferentes puntos que tendrá que seguir y entonces se evitará tener consignas escalones, como son las mostradas en las figuras 5 y 6.

6. Conclusiones

En este artículo se presenta un robot con una arquitectura basada en la integración sistemas multi-agente y control colaborativo. En particular, se ha centrado la atención en el diseño de un agente simple, el agente *goto*, basado en el ajuste difuso de dos controladores de posición.

La colaboración se consigue usando un método basado en lógica difusa, similar a las ideas propuestas en [23]. Esta solución pretende introducir más conocimiento en el proceso de decisión del sistema de control. Los conjuntos difusos modelados en el sistema son dinámicos y se determinan en función de la distancia recorrida por el robot.

Para probar el método, se han desarrollado dos controladores básicos diferentes, uno rápido y el otro preciso, y un controlador colaborativo que combina las dos acciones de control. El resultado del controlador colaborativo se ha integrado como un agente en la arquitectura multiple colaborativa en robot. Seguidamente se han realizado diversos experimentos para comparar la eficiencia de los controladores simples versus el controlador colaborativo.

Los dos controladores, el rápido y el preciso son fáciles de implementar pero muestran comportamientos no deseados cuando funcionan por separado. Al ser la dinámica del robot no lineal, hay condiciones iniciales y finales que no pueden alcanzarse, debido a la restricciones del sistema (no holonómico), con cualquiera de los dos controladores operando por separado. Con el control colaborativo fuzzy, en el cual los dos controladores se combinan, la respuesta del sistema controlado, para diversas consignas, es más rápida que la del controlador más lento y más precisa que la del controlador más rápido. Aún más, llega a funcionar en el caso de algunas de las consignas en las que antes era imposible conseguir con los controladores por separado.

Como trabajo futuro, se prevé extender la solución propuesta al caso de n controladores. También se está explorando la extensión del control colaborativo a otros agentes, como el de *goThrough*, responsable de guiar al robot a través de espacios estrechos como pasillos y puertas.

Referencias

1. Bryson, J. J. (2001). *Intelligence by Desing: Principles of Modularity and Coordination for Engineering Complex Adaptive Agents.* PhD thesis, Massachusetts Institute of Technology.
2. Busquets, D., Sierra, C., and López de Màntaras, R. (2003). A multiagent approach to qualitative landmark-based navigation. *Autonomous Robots*, 15:129 – 154.
3. Campion, G., D'Andrea-Novel, B., and Bastin, G. (1991). Modelling and state feedback control of NonHolonomic mechanical systems. *Proceedings of the 30th Conference on Decision and Control. Brighton, England.*
4. D'Andréa-Novel, B., Bastin, G., and Campion, G. (1991). Modelling and control of non holonomic wheeled mobile robots. *Proceedings of the 1991 IEEE International Conference on Robotics and Automation.*, pages 1130–1135.
5. Dorigo, M. and et al. (2004) Evolving self-organizing behaviours for a swarm-bot. *Autonomous Robots 17* pages: 223-245.

6. Driankov, D., Hellenoorn, H. and Reinfrank, M. (1991). An introduction to Fuzzy Control. *Springer.*.
7. Figueras, A., Colomer, J., and De la Rosa, J. (2002). Supervision of heterogeneous controllers for a mobile robot. In *The XV World Congress IFAC.*
8. Gerkey, B., Mataric, M., and Sukhatme, G. (2002). Exploiting pphysical dynamics for concurrent control of a mobile robot. *Proceedings ICRA '02. IEEE International Conference on Robotics and Automation*, 4:3467 – 3472.
9. Giorgini, P., Kolp, M., and Mylopoulos, J. (2002). Socio-intentional architectures for multi-agent systems: The mobile robot control case. *AOIS-02.*
10. Goldberg, K. and Chen, B. (2001). Collaborative control of robot motion: robustness to error. In *Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 655–660.
11. Innocenti, B., López, B. and Salvi, J.(2006). How MAS support distributed robot control. *ISR/ROBOTIK 2006 - Joint conference on robotics (accepted).*
12. Innocenti, B., Ridao, P., Gascons, N., El-Fakdi, A., López, B. and Salvi, J. (2004). Dynamical model parameters identification of a wheleed mobile robot. *5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles (preprints).*
13. Klir, G. J. and Folger, T. A. (1992). *Fuzzy Sets, Uncertainty, and Information.* Prentice Hall.
14. Hu, H. and Gu, D. (2000). A multi-agent system for cooperative quadruped walking robots. *Proceedings of the IASTED International Conference Robotics and Applications*, pages 1 – 5.
15. Murphy, R. R. (2000). *Introduction to AI Robotics.* The MIT Press.
16. Murray, R., Åström, K., Boyd, S., Brockett, R., and Stein, G. (2003). Future directions in control in an information-rich world. *IEEE Control Systems Magazine*, 23, issue 2:20 – 33.
17. Neves, M. C. and Oliveira, E. (1997). A multi-agent approach for a mobile robot control system. *Proceedings of Workshop on "Multi-Agent Systems:Theory and Applications"(MASTA'97 - EPPIA'97) - Coimbra -Portugal*, pages 1 – 14.
18. Omerdic, E. and Roberts, G. (2004). Thruster fault diagnosis and accommodation for open-frame underwater vehicles. *Engineering Practice*, (12):1575–1598.
19. Ridao, P. and J. Batlle, M. C. (2002). O2CA2, a new object oriented control architecture for autonomy: the reactive layer. *Control Engineering Practice*, 10(8):857–873.
20. Ros, R., Ramon Lopez de Mantaras, Sierra, C., and Arcos, J. L. (2005). A CBR system for autonomous robot navigation. *Frontiers in Artificial Intelligence and Applications 131, IOS Press*, pages 299–306.
21. De la Rosa,J. L. and et al. (1999) Rogi team real: Research on physical agents. *RoboCup-99: Robot Soccer World Cup III* Veloso, Pagello, Kitano (eds), pages 434 –438.
22. Rosenblatt, J. K. (1997). *DAMN: A Distributed Architecture for Mobile Navigation.* PhD thesis, Robotics Institute at Carnegie Mellon University.
23. Saffioti, A. (1997). The uses of fuzzy logic in autonomous robot navigation. *Soft Computing Research Journal*, 1(4):180–197.
24. Spears, W., Spears, D., Hamann, J. and Heil, R. (2004) Distributed, physics based control of swarms of vehicles *Autonomous Robots 17*, pages: 137–162.
25. Soh, L.-K. and Tsatsoulis, C. (2005). A real-time negotiation model and a multi-agent sensor network implementation. *Autonomous Agents and Multi-Agent Systems*, 11(3):215–271.
26. Tounsi, M., Lebret, G., and Gautier, M. (1995). Dynamic control of a nonholonomic mobile robot in cartesian space. *Proceedings of the 34th Conference on Decision and Control*, pages 3825–3830.